

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 6 月 1 8 日
Date of Application:

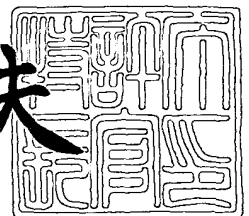
出 願 番 号 特 願 2 0 0 3 - 1 7 2 7 7 0
Application Number:
[ST. 10/C]: [J P 2 0 0 3 - 1 7 2 7 7 0]

出 願 人 株 式 会 社 日 立 製 作 所
Applicant(s):

2 0 0 3 年 7 月 2 5 日

特 許 庁 長 官
Commissioner,
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 K03001911A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 12/00

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

【氏名】 下 蘭 紀夫

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

【氏名】 岩 見 直子

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

【氏名】 本 田 聖志

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社 日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作 田 康夫

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】

要約書 1

【プルーフの要否】

要

【書類名】 明細書

【発明の名称】 ネットワークシステム及びスイッチ

【特許請求の範囲】

【請求項 1】

計算機と、

前記計算機とネットワークを介して接続されるスイッチと、

前記スイッチとネットワークを介して接続される第一の記憶装置と、

前記スイッチとネットワークを介して接続される第二の記憶装置とを有し、

前記スイッチは、あらかじめ前記第一の記憶装置に格納されたデータを前記第二の記憶装置に転送し、

前記計算機は、前記第一の記憶装置に対するデータの読み出し要求を発行し、

前記読み出し要求を受信した前記スイッチは、前記前記第一の記憶装置に対するデータの読み出し要求を前記第二の記憶装置への読み出し要求に変換して、前記第二の記憶装置に変換した前記データの読み出し要求を送信し、

前記第二の記憶装置は、受信した前記データの読み出し要求に対応するデータを前記スイッチに転送し、

データを受信した前記スイッチは、前記受信したデータを前記第一の記憶装置から転送されたデータとして前記計算機に転送することを特徴とするネットワークシステム。

【請求項 2】

前記スイッチに接続される第二の計算機を有し、

前記スイッチは、前記第二の計算機からの指示に基づいて、前記第一の記憶装置に格納されたデータを前記第二の記憶装置へデータに転送することを特徴とする請求項 1 記載のネットワークシステム。

【請求項 3】

前記スイッチは、前記第一の記憶装置に対するデータの読み出し要求を前記第二の記憶装置への読み出し要求に変換する際に、前記データの読み出し要求の送信元を示す情報を別の情報に変換して前記第二の記憶装置に転送し、前記第二の記憶装置から前記変換したデータの読み出し要求に対応するデータを受信した際

には、前記データの送信先として含まれている前記別の情報を前記計算機の情報に変換することを特徴とする請求項2記載のネットワークシステム。

【請求項4】

計算機と、

前記計算機とネットワークを介して接続されるスイッチと、

前記スイッチとネットワークを介して接続される第一の記憶装置と、

前記スイッチとネットワークを介して接続される第二の記憶装置とを有し、

前記スイッチは、あらかじめ前記第一の記憶装置に格納されたデータを前記第二の記憶装置に転送し、

前記スイッチは、前記計算機に前記第一の記憶装置に対応する仮想的な第三の記憶装置を提供し、

前記計算機は、前記第三の記憶装置に対するデータの読み出し要求を発行し、

前記読み出し要求を受信した前記スイッチは、前記前記第三の記憶装置に対するデータの読み出し要求を前記第二の記憶装置への読み出し要求に変換して、前記第二の記憶装置に変換した前記データの読み出し要求を送信し、

前記第二の記憶装置は、受信した前記データの読み出し要求に対応するデータを前記スイッチに転送し、

データを受信した前記スイッチは、前記受信したデータを前記第三の記憶装置から転送されたデータとして前記計算機に転送することを特徴とするネットワークシステム。

【請求項5】

前記仮想的な第三の記憶装置には、前記第二の記憶装置と同じドメインアドレスが割り振られていることを特徴とする請求項4記載のネットワークシステム。

【請求項6】

計算機と、

前記計算機とネットワークを介して接続される第一の記憶装置と、

前記計算機とネットワークを介して接続される第二の記憶装置とを有し、

前記第二の記憶装置は、前記計算機及び前記第一の記憶装置とネットワークを介して接続されるスイッチ部と前記スイッチ部と内部ネットワークを介して接続

される記憶部とを有し、

前記スイッチ部は、あらかじめ前記第一の記憶装置に格納されたデータを前記記憶部に転送し、

前記計算機は、前記第一の記憶装置に対するデータの読み出し要求を発行し、

前記読み出し要求を受信した前記スイッチ部は、前記前記第一の記憶装置に対するデータの読み出し要求を前記記憶部への読み出し要求に変換して、前記記憶部に変換した前記データの読み出し要求を送信し、

前記記憶部は、受信した前記データの読み出し要求に対応するデータを前記スイッチ部に転送し、

データを受信した前記スイッチ部は、前記受信したデータを前記第一の記憶装置から転送されたデータとして前記計算機に転送することを特徴とするネットワークシステム。

【請求項 7】

計算機と、

前記計算機とネットワークを介して接続されるスイッチと、

前記スイッチとネットワークを介して接続される第一の記憶装置と、

前記スイッチとネットワークを介して接続される第二の記憶装置とを有し、

前記計算機は、前記第一の記憶装置に対するデータの読み出し要求を発行し、

前記読み出し要求を受信した前記スイッチは、前記第一の記憶装置に格納されたデータが前記第二の記憶装置に格納されている場合には前記前記第一の記憶装置に対するデータの読み出し要求を前記第二の記憶装置への読み出し要求に変換して、前記第二の記憶装置に変換した前記データの読み出し要求を送信し、前記第一の記憶装置に格納されたデータが前記第二の記憶装置に格納されていない場合には、前記第一の記憶装置へ前記データの読み出し要求を変換せずに送信し、

前記第二の記憶装置は、受信した前記データの読み出し要求に対応するデータを前記スイッチに転送し、

データを受信した前記スイッチは、前記受信したデータを前記第一の記憶装置から転送されたデータとして前記計算機に転送することを特徴とするネットワークシステム。

【請求項 8】

前記スイッチは、前記第一の記憶装置に格納されたデータが前記第二の記憶装置に格納されているかどうかを示す情報を有することを特徴とする請求項 7 記載のネットワークシステム。

【請求項 9】

前記スイッチは、前記第一の記憶装置に格納されたデータが前記第二の記憶装置に格納されていない場合に、前記データの読み出し要求に応じて前記第一の記憶装置から転送されたデータを前記第二の記憶装置へ転送して前記情報を更新することを特徴とする請求項 8 記載のネットワークシステム。

【請求項 1 0】

前記スイッチは、前記第一の記憶装置から転送されたデータを前記第二の記憶装置へ転送する際に、前記第二の記憶装置に前記データを格納するだけの記憶容量が存在しない場合には、前記第二の記憶装置に格納された他のデータのうち、前記計算機に使用される頻度が少ないデータを削除して、前記データを前記第二の記憶装置へ転送し、前記情報を更新することを特徴とする請求項 9 記載のネットワークシステム。

【請求項 1 1】

計算機、第一の記憶装置及び第二の記憶装置と接続されるスイッチであって、外部装置と接続されるポート部と、

前記ポート部が受信したコマンドやデータを振り分ける変換部と、

前記コマンドや前記データをアドレス情報に基づき中継するスイッチ部とを有し、

前記変換部は、前記第一の記憶装置に格納されたデータを前記第二の記憶装置へ予め転送し、前記計算機から受信した前記第一の記憶装置に格納されたデータに対するアクセス要求を前記第二の記憶装置へのアクセス要求に変換し、

前記スイッチ部は、前記第二の記憶装置へのアクセス要求を前記ポートを介して前記第二の記憶装置に送信し、

前記変換部は、前記第二の記憶装置から受信した前記アクセス要求に対応するデータを前記第一の記憶装置から転送されたデータに変換して前記計算機に転送

することを特徴とするスイッチ。

【請求項 1 2】

請求項 1 1 記載のスイッチであって、第二の計算機が接続されており、前記第二の計算機の指示に従って、前記変換部は、前記第一の記憶装置に格納されたデータを前記第二の記憶装置へ転送することを特徴とするスイッチ。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、記憶装置とコンピュータとを接続するネットワークに配置されたスイッチに関する。

【0 0 0 2】

【従来の技術】

企業等で使用される記憶装置の記憶容量の増加に伴って、記憶装置間又は記憶装置とコンピュータとがファイバチャネル等のネットワークを介して接続された一つのシステムが構築されるようになってきた。この記憶装置等を接続するネットワーク又はネットワークによって接続されたシステム全体はストレージエリアネットワーク（以下SAN）と呼ばれる。

【0 0 0 3】

一方、ネットワークでのデータの転送頻度を減らし、コンピュータが記憶装置に格納されたデータへアクセスする時間を短縮する技術としてネットワークキャッシュ技術がある。すなわち、ネットワーク上にデータを一時保存する記憶領域（以下「キャッシュ装置」）を用意し、コンピュータが記憶装置に格納されたデータをキャッシュ装置を介して読み出した場合に、その読み出したデータがキャッシュ装置に格納され、その後のそのデータへのアクセスにはキャッシュ装置が応答を返す。これにより、データへのアクセス時間が短縮される。

【0 0 0 4】

SANにおいてネットワークキャッシュ技術を使用する例が特許文献 1 に開示されている。具体的には、SAN内にメタデータサーバと呼ばれるコンピュータを用意し、記憶装置へアクセスするコンピュータは最初にメタデータサーバにアクセ

スする。メタデータサーバは、アクセスすべきデータの位置をコンピュータに通知する。キャッシュが用いられる場合には、コンピュータにキャッシュを有する装置の位置が通知される。

【0 0 0 5】

また、WWW等のインターネットで、ネットワーク上にキャッシュを有する装置を配置する技術として、トランスペアレントキャッシュと呼ばれる技術がある。本技術では、コンピュータから記憶装置へのデータのアクセス要求（以下「アクセスリクエスト」とも称する）を受取ったスイッチが、そのアクセスリクエストを一度キャッシュを有する計算機（以下「キャッシュサーバ」）に送る。キャッシュサーバが有するキャッシュにアクセスリクエストの対象となるファイルが存在する場合（以下「キャッシュヒット」）、キャッシュサーバは対象となったファイルをアクセスリクエストを発行したコンピュータに送信する。一方、対象となるデータがキャッシュに存在しない場合（以下「キャッシュミス」）は、キャッシュサーバが記憶装置にアクセスリクエストを送信してデータを取得し、取得したデータをアクセスリクエストを発行したコンピュータに送信する。

【0 0 0 6】

【特許文献 1】

特開 2 0 0 2 - 1 3 2 4 5 5 号公報

【0 0 0 7】

【発明が解決しようとする課題】

上述したように、SANにおいてネットワークキャッシュ技術を用いれば、キャッシュヒットの場合には、要求されたデータを取得するまでの時間が短縮される。

【0 0 0 8】

しかし、特許文献 1 に開示された技術では、SAN内にキャッシュ装置を設置できるものの、データのアクセスの為にメタデータサーバを必要とするなど、コンピュータの設定や動作を従来 SAN で使用されているデータアクセスのためのプロトコル、例えば SCSI プロトコルからメタデータサーバを使用するための専用のプロトコルに変更する必要がある。

【0009】

一方、トランスペアレントキャッシュの様にファイル単位でのネットワークキャッシュ技術を用いる場合、キャッシュヒットの判定処理が複雑であるためレスポンスを高速化することが難しい。

【0010】

具体的には、ファイル単位でキャッシュを行う場合、リード要求はファイルの名前などを指定したアクセス要求である。キャッシュサーバは、アクセス要求を受け取ると自身の記憶装置内に保持しているファイルと、そのファイルが記憶装置内のどのブロックに対応するかを記憶するメタデータと呼ばれるデータを読み出し、その中にリード要求が指定するファイルがあるか検索することでキャッシュヒットの判定が行われる。この検索処理は、名前が一致するか等を判定するため、SCSIのようなブロックアクセスのプロトコルがアクセス先の位置を指定するために用いる論理ブロックアドレスという数値の比較に比べて処理が複雑になる。

【0011】

本発明の目的は、SANにおいてコンピュータの設定を変更することなく、データに対するアクセスを高速化し、ネットワーク帯域を節約することである。

【0012】**【課題を解決するための手段】**

上述の目的を達成するために、本発明のネットワークシステムは以下の構成を有する。即ち、計算機と、計算機と接続されるスイッチと、スイッチとネットワークを介して接続される第一の記憶装置と、スイッチとネットワークを介して接続される第二の記憶装置を有するネットワークシステムである。

【0013】

本ネットワークシステムにおいて、スイッチは外部からの指示に基づいて第一の記憶装置に格納されたデータを第二の記憶装置へ転送する。その後、計算機から第一の記憶装置に格納されたデータに対するアクセス要求を受けたスイッチは、そのアクセス要求を第二の記憶装置へのアクセス要求に変換し、変換したアクセス要求を第二の記憶装置へ転送する。その後、第二の記憶装置からデータを受

信したスイッチは、そのデータが第一の記憶装置から送信されたデータとなるように変換して計算機に送信する。

【0014】

尚、スイッチに指示を与えるのはスイッチに接続される第二の計算機でも良い。又、スイッチは、第一の記憶装置に対応する仮想的な記憶装置を計算機に提供しても良い。この場合、計算機は、仮想的な記憶装置へアクセス要求を発行する。

【0015】

更に、本発明の別の態様としては、上述したスイッチと第二の記憶装置が一体となって一つの装置を構成していても良い。

【0016】

又、本発明の別の態様としては、スイッチがあらかじめ第一の記憶装置に格納されたデータを第二の記憶装置へ転送するのではなく、計算機からのアクセス要求に応じて、第一の記憶装置に格納されたデータを第二の記憶装置へ転送しても良い。更に本形態の場合、スイッチは第一の記憶装置に格納されたデータが第二の記憶装置に転送されたかどうかの情報を有し、その情報に基づいてアクセス要求を第一又は第二の記憶装置へ転送しても良い。更に、本形態において、スイッチが第一の記憶装置から第二の記憶装置へデータを転送する際に、第二の記憶装置の記憶容量を確認し、転送するデータを格納する分の記憶容量が第二の記憶装置に存在しない場合には、第二の記憶装置に格納されているデータを一定の基準、例えば計算機における使用頻度、に応じて消去し、そこにデータを転送する構成が考えられる。

【0017】

尚、スイッチの代わりに、第一又は第二の記憶装置がデータの転送を制御しても良い。

【0018】

【発明の実施の形態】

図1は、本発明を適用したコンピュータシステムの第1の実施形態を説明する図である。コンピュータシステムには、SAN101、ホスト105、記憶装置104a及び

記憶装置104bが含まれる。ホスト105、記憶装置104a及び記憶装置104bは、SAN101を介して相互に接続されている。SAN101は、ホスト105、スイッチ102a、102b、102c及び後述するコピー管理スイッチ103を有する。

【0019】

以下、本実施形態においては、ホスト105は記憶装置104aに格納されたデータ（以下「マスタデータ」と称する）に対するリード要求を行うものとする。

【0020】

ホスト105は、CPU1051、メモリ1052及びSAN101へ接続するためのインタフェース1053を有する計算機である。

【0021】

記憶装置104は、データを格納するメディア1043、SAN101へ接続するためのインタフェース1044、ホスト105からの要求に応答するためのプログラムを実行するためのCPU1041及びメモリ1042を有する。尚、記憶装置104が有するメディア1043には、多種の例が考えられる。例えば、メディア1043として、複数のハードディスクドライブから構成されるディスクアレイを採用しても良い。また、記憶装置104は、ホスト105からデータのリード要求を受け取ると、要求に対応するデータをホスト105に送信し、最後に送信を完了したことを示すレスポンスを送信する。

【0022】

スイッチ102a、102b、102cおよび、コピー管理スイッチ103は、相互に接続情報を交換し、後述するルーティング処理に必要なルーティングテーブルを作成する。具体的には、各スイッチ間の通信の負荷（オーバーヘッド）を表す情報（以下「接続コスト」という）を交換する。接続コストは一般にスイッチ間の通信線の通信帯域が小さいほど大きくなる。ただし、管理者等は、後述する管理端末106を介して、接続コストを任意の値に設定することもできる。各スイッチは取得した全ての接続コストから、他の各スイッチへの最も接続コストの和が小さい経路を算出し、それらをルーティングテーブルとして記憶する。

【0023】

スイッチ102a、102b、102c及びコピー管理スイッチ103には、各スイッチに固

有のSANドメインアドレスが割り振られている。また記憶装置104a、104b及びホスト105には、それぞれ固有のSANアドレスが割り振られる。ここで、SANアドレスは、SAN101に接続されるコンピュータ等の装置（以下「ノード」とも言う）と接続されるSAN101内のスイッチのSANドメインアドレス及びSANドメインアドレスで指定される集団（以下「ドメイン」）内部で固有のSANノードアドレスから構成されるアドレスである。

【0024】

各ノードは他のノードへフレームを送受信する時、フレームに送信元と送信先のSANアドレスを付加することで、自分と相手を指定する。スイッチ102a、102b、102c及びコピー管理スイッチ103は、ルーティングテーブル内でフレームの送信先のSANドメインアドレスを検索して、フレームのルーティングを行う。また、フレームの送信先のSANドメインアドレスが当該スイッチ102のSANドメインアドレスと一致している場合は、フレームの送信先のSANノードアドレスと一致するSANノードアドレスを持つ、当該スイッチ102に直接接続するノードへフレームを転送する。尚、フレームとは、SAN101で使用されるプロトコルにおけるデータ又はアクセスリクエストを転送する際の単位である。

【0025】

本実施形態では、接続コストが小さい、距離が短い又は記憶装置104bの方が記憶装置104aよりも高性能などの理由により、ホスト105から記憶装置104aへアクセスするよりも、記憶装置104bへアクセスするほうが、高速にアクセス出来るものとする。

【0026】

コピー管理スイッチ103には、管理端末106が接続されている。本実施形態のコンピュータシステムのユーザ又は管理者は、管理端末106を用いてコピー管理スイッチ103にマスタデータを記憶装置104bへコピーするよう指示する。この際管理者は、管理端末106を介して、マスタデータの位置とコピーされたデータ（以下「コピーデータ」と称する）の位置との対応を示す情報をコピー管理スイッチ103、具体的には、後述するコピー管理テーブル131に登録する。尚、管理者等は、管理端末106を介してホスト105から記憶装置104aへのアクセス頻度などの情報

を収集し、マスタデータ全体ではなくアクセス頻度の高い一部の領域だけを記憶装置104bにコピーする旨、コピー管理スイッチ103に指示しても良い。

【0 0 2 7】

また、スイッチ102a、102b、102c及びコピー管理スイッチ103はネーム管理を行う場合もある。例えばファイバチャネルスイッチは、ノードやポートに割り振られる、世界でユニークな不変の値であるワールドワイドネーム（以下「WWN」）から、SAN101内でユニークなポートのSANアドレスを検索するためのテーブルを有する。WWNには、SAS101に接続するポートに固有なWWPN(World Wide Port Name)と、ノードに固有なWWNN(World Wide Node Name)がある。複数の接続ポートを持つノードは、1つのWWNNと複数のWWPNを持つことができる。また、IPネットワーク上にSCSIプロトコルを流すiSCSIと呼ばれる規格でもiSNSと呼ばれるネーム管理用の体系が存在する。

【0 0 2 8】

図2はコピー管理スイッチ103の構成を示す図である。コピー管理スイッチ103は、他のノードと接続するためのポート111、制御部112、管理端末106と接続するための管理ポート113、アドレス変換部114並びにルーティング処理及びスイッチング処理を行うスイッチング処理部115を有する。スイッチング処理部115は、ルーティング処理を行うためのルーティングテーブルを保持する。

【0 0 2 9】

制御部112は、CPU1121、メモリ1122及び不揮発メモリ1123を有する。アドレス変換部114は、CPU1141及びメモリ1142を有する。尚、本図で開示された構成は実施の一例であり、同等の機能が達成できるならば他の構成であってもよい。

【0 0 3 0】

図3は、コピー管理スイッチ103のメモリ1122、1142及び不揮発メモリ1123に格納されるプログラムやデータを示す図である。

【0 0 3 1】

初期化処理プログラム121は、コピー管理スイッチ103の起動時にCPU1121で実行されるプログラムである。CPU1121は、初期化処理プログラム121を実行する

ことで、他の各プログラムを不揮発メモリ1123からメモリ1122及びメモリ1142に読み出し、後述するコピー管理テーブル131を各アドレス変換部114が有するメモリ1142に読み出す。

【0 0 3 2】

管理端末からの要求処理プログラム122、ルーティングプロトコル処理プログラム123及びネームサービス処理プログラム124は制御部112のメモリ1122に格納され、CPU1121で実行される。アドレス変換処理プログラム126はアドレス変換部114のメモリ1142に格納され、CPU1141で実行される。

【0 0 3 3】

CPU1121は、管理端末からの要求処理プログラム122を実行することで、管理ポート113で受信される管理端末106からの要求に基づき、コピー管理テーブル131の内容を変更する。またCPU1121は、管理端末106からの要求処理プログラム122を実行することで、管理端末106からの要求に基づき、データのコピーを実行する。

【0 0 3 4】

一方、CPU1121は、管理端末からの要求処理プログラム122を実行することで、一般的なスイッチで行われる管理を実行することができる。尚、管理ポート113で用いられるプロトコルは例えばTCP/IPなどがあるが、管理端末106と通信できるならば他のプロトコルが用いられても良い。

【0 0 3 5】

CPU1121は、ルーティングプロトコル処理プログラム123を実行することで、他のスイッチ102とSAN101内における接続に関する情報（以下「接続情報」）を交換することでルーティングテーブルを作成し、スイッチング処理部115が有するメモリに、作成したルーティングテーブルを格納する。

【0 0 3 6】

CPU1121は、ネームサービス処理プログラム124を実行することで、コピー管理スイッチ103に接続されているノードの情報をネームデータベース125に登録する一方、ノードからの検索要求に応答する。ノードからのネームデータベース125への検索要求を受け取るため、制御部112にはSANアドレスが割り振られている。

【 0 0 3 7 】

アドレス変換部114のCPU1141は、アドレス変換処理プログラム126を実行して、ポート111がフレームを受信したことを契機に、コピー管理テーブル131に登録されている情報に基づいて、リード要求・リードデータなどの送信先・送信元SANアドレスを変換する。アドレス変換処理の詳細は後述する。尚、本実施形態ではアドレス変換処理をプログラムに基づいて実行するが、専用ハードウェアでアドレス変換処理を行っても良い。

【 0 0 3 8 】

図4は、コピー管理テーブル131の構造を示す図である。コピー管理テーブル131は、複数のコピー管理エントリ132を有する。各コピー管理エントリ132には、一組のマスタデータとコピーデータの対応関係の情報が保持される。コピー管理エントリ132は、マスタデータを格納する記憶装置104を示すマスタSANアドレスが登録されるフィールド133、マスタデータが格納される記憶装置104内の論理ユニットを示す番号（以下「マスタLUN」）が登録されるフィールド134、論理ユニット内のマスタデータの先頭位置を示す論理ブロックアドレス（以下「マスタLBA」）が登録されるフィールド135、マスタデータのサイズを示すマスタLengthが登録されるフィールド136、コピー管理エントリ132に登録されたマスタデータに対応するコピーデータが格納された記憶装置104を示すコピーSANアドレスが登録されるフィールド137、コピーデータが格納される記憶装置104内の論理ユニットを示す番号（以下「コピーLUN」）が登録されるフィールド138及び論理ユニット内のコピーデータの先頭位置を示す論理ブロックアドレス（以下「コピーLBA」）が登録されるフィールド139を有する。尚、コピーデータのデータ長さはマスタデータと同じなので、コピーLengthを登録するフィールドは不要である。

【 0 0 3 9 】

以下、本発明における各装置の動作の概要について説明する。

【 0 0 4 0 】

本実施形態では、まず、管理者等の指示にしたがって、コピー管理スイッチ103がマスタデータが格納された記憶装置105aから記憶装置105bにマスタデータが転送され、コピーデータが作成される。この際、コピー管理テーブル131にマス

タデータを格納する記憶装置105aとコピーデータを格納する記憶装置105bとの関係を示す情報が登録される。

【0 0 4 1】

上述したコピー処理の終了後、コピー管理スイッチ103は、ホスト105からリード要求を受け取ると、リード要求が含まれるフレームが保持するアドレス情報が上述したマスタデータとコピーデータとの対応関係を示す情報に含まれているか否かを判断することで、リード要求に対応するデータがマスタデータで、かつマスタデータに対応するコピーデータが存在するか判定する。

【0 0 4 2】

リード要求の対象となるマスタデータに対応するコピーデータが存在する場合、コピー管理スイッチ103は、ホスト105から受信した記憶装置104aに格納されたマスタデータへのリード要求を、そのコピーデータを格納する記憶装置104bに対するリード要求に変換する。具体的には、コピー管理スイッチ103は、リード要求に含まれるリクエスト受信先（記憶装置104a）を示すSANアドレスを、コピーデータが格納された記憶装置104bのSANアドレスに変更する。これによりネットワークの効率的な利用を行う。

【0 0 4 3】

以下、コピーデータ作成の手順について説明する。システムの利用者又は管理者は管理端末106を介して、記憶装置104aのSANアドレス、マスタデータの記憶装置104a内部でのアドレスである論理ユニット番号（以下「LUN」）、論理ブロックアドレス（以下「LBA」）、マスタデータのLength、記憶装置104bのSANアドレス、データコピー先のLUN、LBAをコピー管理スイッチ603に送信する。

それらの情報を受信した制御部112は、マスタデータへのリード要求を記憶装置104aに送信する。

【0 0 4 4】

次に、制御部112は、記憶装置104aから送られてきたリードデータをメモリ112に格納する。続けて制御部112は、記憶装置104bにライト要求を送信し、メモリ112に格納されているマスタデータを記憶装置104bに書き込む。このような処理によって、データのコピーが行われる。

【 0 0 4 5 】

尚、上述のコピー動作は処理の一例であり、この方法に限定されるものではない。例えばメモリ1122の代わりにリードデータを格納する専用のバッファを制御部112に設けても良い。また、記憶装置104a自体がコピー処理を行っても良い。例えばSCSIプロトコルの規格には、EXTENDED COPYというコマンドがある。このコマンドを処理することができる記憶装置は、EXTENDED COPYコマンドを受け取ると、コマンドが指定する記憶装置の特定の領域を他の記憶装置の特定の領域にコピーする。

【 0 0 4 6 】

したがって、マスタデータを保持する記憶装置104aがEXTENDED COPYコマンドを処理することができる場合、以下のようにしてもコピー処理を実行できる。

【 0 0 4 7 】

制御部112は、マスタデータを保持する記憶装置104aにコピーを行うようEXTENDED COPYコマンドを送信する。記憶装置104aは、EXTENDED COPYコマンドで指定された記憶領域の内容を、記憶装置104bに転送する。指定された記憶領域に格納されたデータの転送が終了すると、記憶装置104aは、処理の終了をしめすレスポンスをEXTENDED COPYコマンドの送信元、ここではコピー管理スイッチ603へ送信し、コピー処理が完了する。

【 0 0 4 8 】

次に、コピー処理完了後のコピー管理スイッチ103の動作について説明する。上述したように、ホスト105からフレームを受信したコピー管理スイッチ103は、アドレス変換部114でフレームの転送元あるいは転送先のアドレスを変換し、適切な装置へフレームを転送する。

【 0 0 4 9 】

図5は、アドレス変換部114で実行されるアドレス変換処理の手順の一例を示すフローチャートである。

ポート111がフレームを受け取ると、CPU1141は、アドレス変換処理プログラム126の実行を開始する（ステップ151）。CPU1141は、ポート111で受信されたフレームが、ホスト105が発行した記憶装置104へのリード要求を含むフレームである

か否かを判定する（ステップ152）。

【 0 0 5 0 】

フレームがリード要求を含む場合、CPU1141は、リード要求が要求するデータのコピーがコンピュータシステムに存在するか判定する。具体的には、CPU1141は、コピー管理エントリ132に含まれるフィールド133、フィールド134、フィールド135及びフィールド136に登録された情報で示される領域に、リード要求を含むフレームに格納されている送信先SANアドレス、LUN、LBA及びLengthで示される領域が含まれるという条件を満たすコピー管理エントリ132がコピー管理テーブル131に存在するか判定する（ステップ153）。

【 0 0 5 1 】

ステップ153の条件に合致するコピー管理エントリ132が存在する場合、CPU1141は、そのコピー管理エントリ132を用いて、リード要求を含むフレームをコピーデータが格納されている記憶装置104へのリード要求を含むフレームに変換する。具体的には、CPU1141は、リード要求を含むフレームの送信先をフィールド137に登録された値に、LUNをフィールド138に登録された値に、LBAを(LBA+フィールド139に登録された値-フィールド135に登録された値)に変更する。

【 0 0 5 2 】

ステップ152でフレームの内容がリード要求ではないと判断した場合、CPU1141は、フレームの内容がリード要求に対応して記憶装置104から転送されるデータ（以下「リードデータ」）又はレスポンスであるか判定する（ステップ155）。フレームの内容がリードデータ又はレスポンスである場合、CPU1141は、リードデータ又はレスポンスが、コピーデータが格納された記憶装置104bから送信されたものか判定する。具体的には、CPU1141は、フレームの送信元がフィールド137に登録された値と一致するコピー管理エントリ132がコピー管理テーブル131に存在するか判定する（ステップ156）。

【 0 0 5 3 】

フレームがコピーデータが格納された記憶装置104bから送信されていた場合、CPU1141は、ステップ156で見つけたコピー管理エントリ132を用いて、フレームの送信元をフィールド133に登録されたマスタSANアドレスに変更する（ステップ

157)。

【0054】

ステップ154及び157で説明される処理の終了後、CPU1141は、スイッチング処理部115に処理したフレームを送信する。尚、ステップ153及びステップ156で該当するコピー管理エントリ132が存在しない場合並びにステップ155でフレームがリードデータ及びレスポンスではないと判断された場合、CPU1141は、受信したフレームをそのままスイッチング処理部115に送信する（ステップ158）。

【0055】

本実施形態のコンピュータシステムの一連の動作をまとめると、次のようになる。

【0056】

システムの利用者又は管理者が管理端末106を介して、コピー管理スイッチ103にデータコピーを指示し、コピー管理テーブル131に登録する。

【0057】

図4には、スイッチ102cのSANドメインアドレスが4、記憶装置104aのSANノードアドレスが01、コピー管理スイッチ103のSANドメインアドレスが5、記憶装置104bのSANノードアドレスが02であり、長さ100000のマスタデータが記憶装置104aのLUN0のLBA0から格納されていて、マスタデータ全体を記憶装置104bのLUN5のLBA50000から始まる領域にコピーした場合のマスタデータとコピーデータの対応関係を示す情報がコピー管理エントリ132の各フィールドに登録された例が示されている。

【0058】

ホスト105が記憶装置104aへのリード要求を行う。リード要求に対応するフレームには、送信先である記憶装置104aのSANアドレス、LUN、LBA、Length及び送信元であるホスト105のSANアドレスが含まれる。

【0059】

リード要求は送信先のSANアドレスに基づいてルーティングされ、スイッチ102aを経由してコピー管理スイッチ103に届く。リード要求を受取ったコピー管理スイッチ103は、コピー管理テーブル131の情報とフレームの内容とを照合する。リ

ード要求に対応するコピーデータが存在する場合、リード要求をコピーデータへのリード要求に変換し、記憶装置104bに変換されたフレームをルーティングする。

【 0 0 6 0 】

リード要求を受取った記憶装置104bは、リード要求の対象であるコピーデータを読み出し、リード要求の送信元であるホスト105へリードデータを送信する。リードデータのフレームには、送信先であるホスト105のSANアドレス及び送信元である記憶装置104bのSANアドレスが含まれる。

【 0 0 6 1 】

リードデータを受取ったコピー管理スイッチ103は、コピー管理テーブル131の情報に基づきリードデータの送信元を記憶装置104aに変更する。その後、リードデータは送信先のSANアドレスに基づいてルーティングされ、スイッチ102aを通りホスト105に届く。ホスト105は、リードデータを記憶装置104aから送信されたものとして受け取る。

【 0 0 6 2 】

一連の動作により、記憶装置104aへのリード要求は、実際にはSAN101でより高速にアクセスできる記憶装置104bで処理されるので、レスポンスが高速になり、スイッチ102b、102cおよび記憶装置104aにかかる負荷を減らすことができる。

【 0 0 6 3 】

図 6 は、本発明を適用したコンピュータシステムの第二の実施形態を示す図である。第一の実施形態とは、コピー管理スイッチ203が複数存在する点が異なっている。尚、他の構成については第一の実施形態と同一であるので、説明を省略する。

【 0 0 6 4 】

また、本実施形態では、ホスト105から記憶装置104aへの最短経路は、ホスト105→スイッチ102a→コピー管理スイッチ203a→スイッチ102b→記憶装置104aとし、ホスト105から記憶装置104bへの最短経路はホスト105→スイッチ102a→コピー管理スイッチ203b→記憶装置104bとする。またホスト105から記憶装置104aの接続コストよりホスト105から記憶装置104bの接続コストの方が小さいものとする

。

【 0 0 6 5 】

システムの管理者又は利用者は、コピー管理スイッチ203a及び203bと接続される管理端末106を介して、記憶装置104aが保持するマスタデータを記憶装置104bにコピーする指示をコピー管理スイッチ203bに行い、マスタデータとコピーデータの対応を示す情報を、コピー管理スイッチ203a、203bが有するコピー管理テーブル231に登録する。

【 0 0 6 6 】

また、コピー管理スイッチ203が有するメモリ1122には、第一の実施形態におけるプログラム等に加え、代理アドレステーブル241が格納される。さらに、CPU 1141が実行するアドレス変換処理プログラム226の内容も第一の実施形態と異なる。

【 0 0 6 7 】

また、不揮発メモリ1123に格納されるコピー管理テーブル231が有するコピー管理エントリ232は、第一の実施形態のコピー管理エントリ132の各フィールドに加え、ローカルフラグに登録するフィールド240を有する。

【 0 0 6 8 】

ローカルフラグは、コピー管理スイッチ203が有するコピー管理エントリ232のマスタSANアドレス133に対応する記憶装置104と、当該コピー管理スイッチ203を含めた複数のコピー管理スイッチ203との間の接続関係を示すフラグである。具体的には、複数のコピー管理スイッチ203と記憶装置104との間に存在する機器の個数に応じて、値が設定される。以下、機器の個数が少ないことを「近い」と表現する。

【 0 0 6 9 】

本実施形態では、コピー管理スイッチ203aよりコピー管理スイッチ203bの方が記憶装置104bの近くに接続されているので、管理端末106はコピー管理スイッチ203aにはローカルフラグを0に、コピー管理スイッチ203bにはローカルフラグを1にしたコピー管理エントリ232をコピー管理テーブル231に登録する。

【 0 0 7 0 】

図 7 は、代理アドレステーブル 241 の構成を示す図である。代理アドレステーブル 241 は、データを要求するホスト 105 を示す SAN アドレス（以下「ホスト SAN アドレス」）、マスタ SAN アドレス、及び本実施形態におけるアドレス変換処理で用いられる代理アドレス（以下「代理 SAN アドレス」）との対応関係を記録するためのテーブルである。

【 0 0 7 1 】

代理アドレステーブル 241 は複数の代理アドレスエントリ 242 を有する。各代理アドレスエントリ 242 は、ホスト SAN アドレスが登録されるフィールド 243、マスタ SAN アドレスが登録されるフィールド 244 及び代理 SAN アドレスが登録されるフィールド 245 を有する。代理アドレステーブル 241 の使用方法については後述する。

【 0 0 7 2 】

図 8 は、本実施形態における CPU1141 のアドレス変換処理の手順の一例を示すフローチャートである。

【 0 0 7 3 】

ポート 111 がフレームを受信したら、CPU1141 は、アドレス変換処理プログラム 226 の実行を開始する（ステップ 251）。CPU1141 は、受信したフレームの内容がリード要求か否かを判定する（ステップ 252）。受信したフレームの内容がリード要求である場合、CPU1141 は、リード要求が要求するデータのコピーがコンピュータシステム内に存在するか判定する。具体的には、CPU1141 は、コピー管理エントリ 132 に含まれるフィールド 133、フィールド 134、フィールド 135 及びフィールド 136 に登録された情報で示される領域に、リード要求を含むフレームに格納されている送信先 SAN アドレス、LUN、LBA 及び Length で示される領域が含まれるという条件を満たすコピー管理エントリ 132 がコピー管理テーブル 131 に存在するか判定する（ステップ 253）。

ステップ 253 で示される条件に合致するコピー管理エントリ 232 が存在する場合、CPU1141 は、ステップ 253 で見つけたコピー管理エントリ 232 を用いて、リード要求をコピーデータが格納される記憶装置 104 へのリード要求に変換する。具体的には、リード要求を含むフレームの送信先をフィールド 137 に登録されたコピ

ーSANアドレスに、フレームのLUNをフィールド138に登録されたコピーLUNに、フレームのLBAを(LBA+フィールド139に登録されたコピーLBAの値-フィールド135に登録されたマスタLBAの値)に変更する(ステップ254)。

【0074】

ステップ253でフレームの送信先がコピー管理テーブル231に含まれていない場合及びフレームをコピーデータが格納されている記憶装置104へのリード要求に変換した場合、CPU1141は、リード要求がCPU1141が属するコピー管理スイッチ203に接続された記憶装置104へのものかどうか判定する。具体的には、まずリード要求の送信先のSANアドレス、LUN、LBA及びLengthで表される領域を、コピーSANアドレス137、コピーLUN138、コピーLBA139とマスタLength136で表される領域が含み、かつローカルフラグ240が1であるコピー管理エントリ232がコピー管理テーブル231に存在するか判定する(ステップ255)。

【0075】

ローカルフラグが1であるコピー管理エントリ232がある場合、CPU1141は、リード要求を含むフレームの送信元を変更する。これは、コピーデータへのリード要求と、コピーデータを保持する記憶装置104が保持するコピーデータ以外のデータへのリード要求とを区別できるようにするためである。具体的には、まずCPU1141は代理SANアドレスを生成する。代理SANアドレスは、当該コピー管理スイッチ203に割り振られたSANDメインアドレスを含み、他のいかなるノードのSANアドレスや、代理アドレステーブル241に保持される代理アドレスエントリ242のフィールド245に登録された代理SANアドレスと重複しないように決定される。

【0076】

次に、CPU1141は、代理アドレステーブル241内の使用されていない代理アドレスエントリ242に、リード要求を発行したホスト105に対応するホストSANアドレス、マスタSANアドレス及び代理アドレスとの対応関係を登録する。具体的には、ステップ255で見つかったコピー管理エントリ232を用いて、CPU1141は、リード要求の送信元であるホスト105を示すホストSANアドレスをフィールド243に、マスタSANアドレス133をフィールド244に、生成した代理SANアドレスをフィールド245に登録する。その後、CPU1141は、リード要求であるフレームの送信元を、

生成した代理アドレスに変更する（ステップ256）。

【 0 0 7 7 】

ステップ252で受信したフレームの内容がリード要求でないと判断した場合、CPU1141は、受信したフレームの内容がリードデータであるか判定する（ステップ257）。受信したフレームの内容がリードデータである場合、CPU1141は、リードデータの送信先が当該コピー管理スイッチ203の生成した代理SANアドレスであるか判定する。具体的には、CPU1141は、リードデータの送信先を示すSANアドレスがフィールド245に登録された代理SANアドレスと一致する代理アドレスエントリ242が代理アドレステーブル241に存在するか判定する（ステップ258）。

【 0 0 7 8 】

条件を満たす代理アドレスエントリ242が見つかった場合、CPU1141は、ステップ258で見つけた代理アドレスエントリ242に登録された情報を用いて、フレームの送信元をフィールド244に登録されたマスタSANアドレスに、送信先をフィールド243に登録されたホストSANアドレスに変更する（ステップ259）。

【 0 0 7 9 】

ステップ257で受信したフレームの内容がリードデータでないと判断した場合、CPU1141は、受信したフレームの内容がレスポンスであるか判定する（ステップ260）。フレームの内容がレスポンスである場合、CPU1141は、フレームの送信先が当該コピー管理スイッチ203の生成した代理SANアドレスで示されるノードであるか判定する。具体的には、CPU1141は、フレームの送信先を示すSANアドレスがフィールド245に登録された代理SANアドレスと一致する代理アドレスエントリ242が代理アドレステーブル241に存在するか判定する（ステップ261）。

【 0 0 8 0 】

ステップ261で条件を満たす代理アドレスエントリ242が見つかった場合、CPU1141は、見つかった代理アドレスエントリ242に登録された情報を用いて、フレームの送信元をフィールド244に登録されたマスタSANアドレスに、送信先をフィールド243に登録されたホストSANアドレスに変更する。また、CPU1141は、代理アドレスエントリ242を代理アドレステーブル241から消去する（ステップ262）。

【 0 0 8 1 】

フレームの送信先がステップ255で示される条件を満たさない場合、ステップ258及び261でフレームの送信先が代理SANアドレスでない場合並びにステップ256、259及び262でフレームのアドレス変換が済んだ場合、CPU1141は、スイッチング処理部115へ処理したフレームを送信する（ステップ263）。

【 0 0 8 2 】

本実施形態におけるコンピュータシステムの一連の動作をまとめると、次のようになる。

【 0 0 8 3 】

システムの利用者又は管理者が管理端末106を介して、コピー管理スイッチ103にデータコピーを指示し、各コピー管理スイッチのコピー管理テーブル231へその情報を登録する。

【 0 0 8 4 】

ホスト105が記憶装置104aへのリード要求を行う。リード要求は送信先である記憶装置104aのSANアドレス、LUN、LBA、Lengthと送信元であるホスト105のSANアドレスを含む。リード要求は送信先のSANアドレスに基づいてルーティングされ、スイッチ102aを経由してコピー管理スイッチ203aに届く。

【 0 0 8 5 】

コピー管理スイッチ203aはリード要求を受け取ると、コピー管理テーブル231の情報と照合する。コピー管理スイッチ203aは、リード要求の送信先を記憶装置104bが保持するコピーデータへのリード要求に変換しルーティングする。ただし、記憶装置104bはコピー管理スイッチ203aに接続されていないので、代理アドレスは生成されず、送信元も変更されない。

【 0 0 8 6 】

書き換えられたリード要求は送信先のSANアドレスに基づいてルーティングされ、コピー管理スイッチ203bに届く。リード要求を受取ったコピー管理スイッチ203bは、受取ったフレームの内容をコピー管理テーブル231の情報と照合する。その結果、コピー管理スイッチ203bは、代理SANアドレスを生成し、ホスト105のSANアドレス、マスタである記憶装置104aのSANアドレス及び代理SANアドレスの対応関係を代理アドレステーブル241の代理アドレスエントリ242に記録する。そ

の後、コピー管理スイッチ203bは、リード要求の送信元を代理SANアドレスに変更してフレームをルーティングする。

【0087】

リード要求を受取った記憶装置104bは、リード要求に対応するコピーデータを読み出し、送信先を代理SANアドレスにしてリードデータをコピー管理スイッチ203bへ送信する。リードデータを含むフレームには、送信先である代理SANアドレスと送信元である記憶装置104bのSANアドレスが含まれる。

【0088】

リードデータのフレームがコピー管理スイッチ203bに届くと、コピー管理スイッチ203bは代理アドレステーブル241の情報に基づき、リードデータの送信元を記憶装置104aのSANアドレス、送信先をホスト105のSANアドレスに変更する。その後、リードデータを含むフレームは送信先のSANアドレスに基づいてルーティングされ、スイッチ102aを通りホスト105に届く。ホスト105はリードデータを記憶装置104aから送信されたものとして受け取る。

【0089】

一方、記憶装置104bからのレスポンスがコピー管理スイッチ203bに届くと、コピー管理スイッチ203bは、送信元を記憶装置104aのSANアドレス、送信先をホスト105のSANアドレスに変更しルーティングすると同時に、代理アドレステーブル241から対応関係を記録している代理アドレスエントリ242を消去する。ホスト105はレスポンスを記憶装置104aから送信されたレスポンスとして受け取る。

【0090】

本実施形態では、ホスト105とマスタデータを保持する記憶装置104aの経路上にコピー管理スイッチ203aが存在している。しかし第1の実施形態と異なり、コピー管理スイッチ203aにはコピーデータを保持する記憶装置104bは接続されていない。それでも、本実施形態では、リード要求はコピーデータを保持する記憶装置104bへのリード要求へ変換され、コピーデータを保持する記憶装置104bが接続しているコピー管理スイッチ203bに届き、コピーデータが記憶装置104aからのリードデータとしてホスト105に送信される。

【0091】

また、本実施形態では代理SANアドレスを用いることで、コピー管理スイッチ203によって変更されたリード要求と、元々コピーを保持する記憶装置104bへ発行されたリード要求などのコマンドとを区別することが可能である。そのため、コピーデータを有する記憶装置104bを通常の記憶装置104bとして用いることが可能になる。さらに、リード要求ごとに異なる代理SANアドレスを用いるので、複数の記憶装置104に保持されるマスタデータを一つの記憶装置104bにコピーして、コピーデータとして用いることが可能になる。

【0092】

尚、本実施形態では代理SANアドレスを用いて、記憶装置104からのリードデータを、そのアドレス情報を変換する必要のあるリードデータと、変換する必要のないリードデータを分類した。しかし、フレームにリード要求とリードデータとレスポンスの対応を認識できるような情報が付加されている場合は代理SANアドレスを用いず、その付加情報でリードデータを分類することもできる。例えば、ファイバチャネルプロトコルではエクスチェンジIDと呼ばれるIDが各フレームに付加されているので、この情報に基づいてリードデータ、レスポンスを分類することも可能である。

【0093】

図9は、本発明の第3の実施形態を適用したコンピュータシステムの構成例を示す図である。SAN101は、スイッチ102a、102b及びコピー管理スイッチ303を有する。また、ホスト105、記憶装置104a及び記憶装置104bがSAN101に接続されている。記憶装置104aにはマスタデータが格納されている。

【0094】

尚、本実施形態においても、ホスト105と記憶装置104aとの間の通信線の接続コストよりホスト105と記憶装置104bとの間の通信線の接続コストの方が小さいものとする。

【0095】

システムの使用者や管理者は、コピー管理スイッチ303に接続された管理端末106を介して、上述した他の実施形態と同様に、マスタデータのコピーを行い、マスタデータとコピーデータの対応関係の情報を、コピー管理スイッチ303のコピ

ー管理テーブル231に記録し管理する。本実施例では、これらに加えて後述する仮想アドレステーブル341も記録し、管理する。

【0096】

コピー管理スイッチ303は、後述する方法を用いて、コピー管理スイッチ303に接続される装置に対して、仮想的な記憶装置104（以下「仮想記憶装置307」）が存在するかのように振舞う。

【0097】

以下、本実施形態においては、ホスト105はマスタデータが仮想記憶装置307に格納されていると判断して、仮想記憶装置307へリード要求を行う。これは、後述するコピー管理エントリ232a、232bが管理端末106によってコピー管理スイッチ303に登録されているため、コピー管理スイッチ303が仮想記憶装置307へのリード要求を、コピーデータの有無により記憶装置104aないし記憶装置104bへのリード要求に変換するからである。これによりネットワークの効率的な利用を行う。尚、ホスト105には仮想記憶装置307のWWNが設定されている。ホスト105はネームサービスを使って、仮想記憶装置307のWWNから仮想記憶装置307のSANアドレス（2つ）を知る。

【0098】

尚、本実施形態におけるコピー管理スイッチ303の構成は第1の実施形態のコピー管理スイッチ103と同様である。しかし、コピー管理スイッチ303が有するメモリに登録される情報等については、第二の実施形態と比較して、下記に示すような違いがある。

【0099】

第一に、CPU1121が初期化処理プログラム321を実行して、後述する仮想アドレステーブル341を不揮発メモリ1123から読み出し、仮想アドレステーブル341に記録されている仮想記憶装置307のアドレス情報をネームデータベース125に登録する点である。第二に、CPU1121が端末からの要求処理プログラム322を実行して、第2の実施形態における処理に加え、管理ポート113で受信される管理端末106からの要求に答え、不揮発メモリ1123に保持される仮想アドレステーブル341の変更を行う点である。

【 0 1 0 0 】

図 1 0 は、仮想アドレステーブル341の内容を示す図である。仮想アドレステーブル341は、複数の仮想アドレスエントリ342を有する。仮想アドレスエントリ342は1つの仮想的なノード、例えば仮想記憶装置307に対応し、仮想的なノードのSANアドレス（以下「仮想SANアドレス」）及び仮想的なWWN（以下、「仮想WWPN」、「仮想WWNN」と称する）が、仮想アドレスエントリ342のエントリ343、344及び345に登録される。

【 0 1 0 1 】

尚、本実施形態におけるコピー管理テーブル231、代理アドレステーブル241は第2の実施形態と同様のものを用いる。アドレス変換処理プログラム226も第2の実施形態と同様である。

【 0 1 0 2 】

本実施形態におけるコンピュータシステムの一連の動作をまとめると、次のようになる。

【 0 1 0 3 】

はじめに、システムの利用者や管理者が、コピー管理スイッチ303に接続された管理端末106を介して、記憶装置104aに格納されたマスタデータを記憶装置104bコピーする指示をコピー管理スイッチ303に行う。続けて、仮想アドレステーブル341とコピー管理テーブル231を設定する。スイッチ102aのSANドメインアドレスが4、コピー管理スイッチのSANドメインアドレスが5、記憶装置104aのSANアドレスが401、記憶装置104bのSANアドレスが501、仮想記憶装置307に割り当てるSANアドレスを502、WWPNを1234、WWNNを5678とし、マスタデータが記憶装置104aのLUN0、LBA0から始まる長さ100000の領域に格納されており、マスタデータの先頭から長さ50000の部分を記憶装置104bのLUN0、LBA0から始まる領域にコピーした場合のコピー管理テーブル231、コピー管理エントリ232a、232bと仮想アドレステーブル341、仮想アドレスエントリ342にセットされる値を図10、図11に示す。コピー管理エントリ232aは、仮想記憶装置307のLUN0、LBA0から始まる長さ50000の部分へのリード要求が、記憶装置104bへのリード要求に変換されることを示す。一方コピー管理エントリ232bは、仮想記憶装置307のLUN0、LBA50000から始

まる長さ50000の部分へのリード要求が、記憶装置104aへのリード要求に変換されることを示す。

【0104】

設定終了後、ホスト105が仮想記憶装置307へのリード要求を発行する。リード要求のフレームには、送信先である仮想記憶装置307のSANアドレス、LUN、LBA、Length及び送信元であるホスト105のSANアドレスが含まれる。リード要求のフレームは、送信先のSANアドレスに基づいてルーティングされ、スイッチ102aを経由してコピー管理スイッチ303に届く。

【0105】

リード要求を受取ったコピー管理スイッチ303は、リード要求のフレームに含まれている情報をコピー管理テーブル231の情報と照合する。照合の結果、リード要求がコピー管理エントリ232aにヒットしたなら、コピー管理スイッチ303は、第二の実施形態と同様に、リード要求の送信先をコピーデータを有する記憶装置104bへのリード要求に変換し、代理アドレスを生成し、送信元を代理アドレスに変更して、そのリード要求のフレームを記憶装置104bに転送する。そして、コピー管理スイッチ303bは、ホスト105a、仮想記憶装置307及び代理アドレスの対応を代理アドレステーブル241の代理アドレスエントリ242に記録する。また、照合の結果、リード要求がコピー管理エントリ232bにヒットした場合は、リード要求を104aに格納されたマスターデータへのリード要求に変換し、同様の処理を行う。

【0106】

リード要求を受取った記憶装置104bは、指定されたデータを読み出し、送信先を代理SANアドレスに設定してリードデータを送信する。リードデータのフレームには送信先を示す代理SANアドレス及び送信元である記憶装置104bのSANアドレスが含まれる。

【0107】

リードデータのフレームを受信すると、コピー管理スイッチ303は代理アドレステーブル241の情報に基づき、リードデータのフレームの送信元を仮想記憶装置307のSANアドレス、送信先をホスト105のSANアドレスに変更する。リードデー

タは送信先のSANアドレスに基づいてルーティングされ、スイッチ102aを通りホスト105に届く。ホスト105は、リードデータを仮想記憶装置307から送信されたものとして受け取る。

【0 1 0 8】

第1、第2の実施形態では、ホスト105とマスタデータを保持する記憶装置104aとの間のネットワークの経路（以下「パス」）上にコピー管理スイッチ103等があり、そのコピー管理スイッチ103等でリード要求のフレームが変更されていた。しかし本実施形態では、コピー管理スイッチ303が仮想記憶装置307を提供するため、ホスト105と記憶装置104aとの間のパス上には無いコピー管理スイッチ303に直接リード要求が到達する。

【0 1 0 9】

また、本実施形態の一つとして、図12のような構成を取ることもできる。この構成では、記憶装置104aがコピー管理スイッチ303aに接続される。管理者等は、管理端末106を介してコピー管理スイッチ303a、303bのコピー管理テーブル231や仮想アドレステーブル341を設定し、それぞれのコピー管理スイッチが仮想記憶装置307を提供するようにする。

【0 1 1 0】

この場合、ホスト105a、105bが各コピー管理スイッチ303のネームデータベースを参照したときに、仮想記憶装置307が複数のポートを持つノードとして認識されるように、コピー管理スイッチ303a、303bの仮想アドレステーブル341において、WWNNが等しい仮想アドレスエントリ342が登録される。

【0 1 1 1】

ホスト105a、105bでは、アクセスすべき記憶装置104がWWNNで設定される。ホスト105a、105bはネームデータベースを参照し、アクセスすべき記憶装置104（実際には仮想化記憶装置307）の二つのSANアドレスを取得する。この場合、ホスト105は、どちらのSANアドレスを使用しても、仮想化記憶装置307にアクセスすることができる。これら複数のSANアドレスで示される複数のポートから一つを選択する方法としては、以下の二つが考えられる。

【0 1 1 2】

一つは、ホスト105がスイッチ102ないしコピー管理スイッチ303からSAN101のトポロジ情報を取得できる場合、より接続コストの少ないポートを選択し、そのポートへリード要求を送信する。

【0 1 1 3】

もう一つは、ホスト105がトポロジ情報が取得できない場合、ホスト105が両方のポートにリード要求を送信し、より速くアクセスできるポートを選択するという方法である。

【0 1 1 4】

尚、上述の構成においては、マスタデータ全体を記憶装置104bにコピーしている際に記憶装置104bが故障した場合、コピー管理スイッチ303が記憶装置104の故障を検出できると仮定すると、コピー管理エントリ232を変更することで、無事な記憶装置104aへアクセス要求のルーティングが行われるようなフェイルセーフを行うことが可能である。

【0 1 1 5】

コピー管理スイッチ303は、記憶装置104の故障をいくつかの方法で検出できる。例えば光ファイバを通信線として用いている場合、物理的な接続が切れることは、ポート111で光が消える形で検出できる。また記憶装置104bからのレスポンスにはエラー情報が含まれるため、ポート111においてレスポンスの内容を監視することによって、コピー管理スイッチ303が記憶装置104の故障を検出することも可能である。以下、フェイルセーフの例を説明する。

【0 1 1 6】

記憶装置104bの故障を検出したコピー管理スイッチ303は、管理端末106に故障の発生を通知し、ユーザ等が管理端末106を介してコピー管理スイッチ303のコピー管理テーブル231を設定しなおす。例えば、図9の記憶装置104aが故障した場合、ユーザは、管理端末106を介してコピー管理スイッチ303のコピー管理エントリ232aを削除し、コピー管理エントリ232bのマスタLBA135bとコピーLBA139bを0にし、マスタLength136bを100000にする。

【0 1 1 7】

これにより、コピー管理スイッチ303は、ホスト105から仮想記憶装置307への

リード要求を全て記憶装置104aにルーティングするようになる。第一及び第二の実施例でも同様のフェイルオーバー処理が可能である。

【0 1 1 8】

また、図 1 2 の構成の場合は、送信したリード要求の結果がタイムアウトするか否かによって、ホスト105が選択するポートを切替えることでフェイルオーバーを実現することができる。

【0 1 1 9】

図 1 3 は、本発明の第 4 の実施形態を適用したコンピュータシステムの構成例を示す図である。本実施形態は、コピー管理スイッチ403a及び403bが、仮想スイッチ408を提供する点が第 3 の実施形態とは異なる。

【0 1 2 0】

本実施形態におけるコピー管理スイッチ403の構成は第 3 の実施形態におけるコピー管理スイッチ103と同様であるが、コピー管理スイッチ403が有するメモリ1122に登録される情報等が、以下の点で第 3 の実施形態とは異なる。

【0 1 2 1】

第一に、仮想アドレステーブル441に、第 3 の実施形態の仮想アドレステーブル341に加えて仮想ドメインアドレスを登録するエントリ446が追加される。エントリ446に登録される仮想ドメインアドレスは、仮想スイッチ408のSANドメインアドレスを示す。

【0 1 2 2】

第二に、CPU1121がルーティングプロトコル処理プログラム422を実行して、エントリ446に登録された仮想ドメインアドレスで指定されるSANドメインアドレスを持つ仮想スイッチ408が接続されているという情報を他のスイッチと交換し、ルーティングテーブルを作成する点である。このとき、ルーティングの整合性を保つため、コピー管理スイッチ403aと仮想スイッチ408との間の接続コストとコピー管理スイッチ403bと仮想スイッチ408との間の接続コストは等しいように設定する。

【0 1 2 3】

以下、本実施形態の処理の一連の流れを説明する。

【 0 1 2 4 】

はじめに、システムの利用者や管理者が、コピー管理スイッチ403a、403bに接続された管理端末106を介して、記憶装置104aに格納されたマスターデータを記憶装置104bにコピーする指示をコピー管理スイッチ403a（または403b）に行う。続けて、管理者等は、各コピー管理スイッチ403に登録された仮想アドレステーブル441とコピー管理テーブル231に情報を設定する。

【 0 1 2 5 】

ここで、本実施形態では、コピー管理スイッチ403aのSANドメインアドレスが4、コピー管理スイッチ403bのSANドメインアドレスが5、仮想スイッチ408のSANドメインアドレスを8、記憶装置104aのSANアドレスが401、記憶装置104bのSANアドレスが501、仮想記憶装置307に割り当てるSANアドレスを801、WWPNを1234、WWNNを5678とし、マスターデータが記憶装置104aのLUN0、LBA0から始まる長さ100000の領域に格納されており、マスターデータの先頭から長さ50000の部分を記憶装置104bのLUN0、LBA0から始まる領域にコピーするものとする。

【 0 1 2 6 】

また、ホスト105と各コピー管理スイッチ403との間の接続コストは以下のとおりであると仮定する。すなわち、ホスト105aに関しては、コピー管理スイッチ403bとの間の方がコピー管理スイッチ403aとの間より接続コストが小さく、ホスト105bに関しては、コピー管理スイッチ403aとの間の方がコピー管理スイッチ403bとの間より接続コストが小さい。

【 0 1 2 7 】

ホスト105aが仮想記憶装置307にリード要求を発行すると、接続コストが最小となるルーティングによってリード要求はコピー管理スイッチ403bに到達し、その後は第3の実施形態と同様の処理が行われる。

【 0 1 2 8 】

ホスト105bが仮想記憶装置307にリード要求を発行すると、リード要求はコピー管理スイッチ403aに到達し、第3の実施形態と同様の処理が行われる。

本実施形態により、ホスト105は、図12の構成のようにパスを選択しなくても、接続コストが小さいコピー管理スイッチ403にリード要求を送ることができ

る。これは、コピー管理スイッチ403a, 403bと仮想スイッチ308の間の接続コストが等しいので、ホスト105から仮想スイッチ308のSANドメイン宛に送信されたフレームはホスト105の直近のコピー管理スイッチ403aないし403bにたどり着くからである。

【0 1 2 9】

図14は、本発明の第5の実施形態を適用したコンピュータシステムの構成例を示す図である。本実施形態は、コピー管理スイッチ503が記憶装置104c及び104dを有する点が第一の実施形態と異なる。

【0 1 3 0】

本実施形態においては、記憶装置104aがマスタデータを保持しており、ホスト105はマスタデータへリード要求を行う。また、本実施形態ではコピー管理スイッチ503がコピーデータを保持する。そして、ホスト105からマスタデータへのリード要求を受取ったコピー管理スイッチ503は、内部の記憶装置104c、104dに保持しているコピーデータを読み出して、ホスト105にリードデータを送信する。

【0 1 3 1】

図15は、コピー管理スイッチ503の内部構成の具体例を示す図である。

コピー管理スイッチ503は、複数のポート5031及びポートプロセッサ5033を有するプロトコル変換部5032、ディスク制御部5035、ハードディスク5036、管理部5037及びこれらの構成部分を接続するスイッチ部5034を有する。ポートプロセッサ5033はCPU及びメモリを有し、管理部5037はCPU、メモリ及び記憶装置を有する。

【0 1 3 2】

ディスク制御部5035は、ディスク制御部5035に接続された複数のハードディスクの記憶容量を一つ又は複数の論理的な記憶領域として他の装置に提供する。コピー管理スイッチ503は、自スイッチ503に含まれる複数のディスク制御部5035が提供する論理的な記憶領域をまとめて一つ又は複数の仮想的な記憶領域として、コピー管理スイッチ503に接続される装置に提供する。

【0 1 3 3】

コピー管理スイッチ503は、ポート5031を介して、他のノードとコマンドやデ

ータを送受信する。ポート5031を介してコマンド等を受信したプロトコル変換部5032は、受信したコマンドやデータのプロトコルを変換してスイッチ部5034に転送する。ここで、プロトコル変換部5032は、受信したコマンドがコピー管理スイッチ503の提供する仮想的な記憶領域に対するものか判別する。受信したコマンドが自記憶領域に対するものである場合、プロトコル変換部5032は、その記憶領域に対応するディスク制御部5035へコマンドを発行する。自記憶領域に対応しない場合、プロトコル変換部5032は、受け取ったフレームをそのままスイッチ部5034に転送する。

【0134】

コピー管理スイッチ503の提供する仮想的な記憶領域とディスク制御部5035が提供する論理的な記憶領域との対応関係を示す情報は管理部5037の記憶装置に格納される。またこの情報は、コピー管理スイッチ503の起動時に管理部5037のCPUによってポートプロセッサ5033が有するメモリに格納される。また、管理部5037は内部のメモリにネームデータベースを保持し、WWNN等の問い合わせに応答する処理を内部CPUにて実行する。

スイッチ部5034は、フレームのアドレス情報に基づきルーティング処理を行う。

【0135】

ディスク制御部5035は、提供する論理的な記憶領域と接続されている各ハードディスク5036が有する記憶領域との対応関係の情報を内部のメモリに保持している。ディスク制御部5035は、スイッチ部5034からコマンド（ここでは、フレームの内部にコマンドが格納されている）を受け取ると、メモリに保持された対応関係の情報をを用いて、コマンドが指定する記憶領域に対応するハードディスク5036と、その内部での格納位置を決定する。そして対応するハードディスク5036へデータの読み出し等の命令を発行し、コマンドを処理する。

【0136】

尚、図15においてスイッチ部5034、ディスク管理部5035及び管理部5037が各々2個ずつ存在するのは、冗長性を持たせて信頼性を向上させるためであり、本発明においてこの構成が必須というわけでは無い。

【 0 1 3 7 】

本実施形態のコピー管理スイッチ503において、プロトコル変換部5032のポートプロセッサ5033が有するメモリにアドレス変換処理プログラム126が格納される。アドレス変換処理プログラム126は、プロトコル変換部5032のポート5031がフレームを受信した場合、またフレームをスイッチ部5034から受信した場合にポートプロセッサ5033aで実行される。また、管理部5037内部のメモリには、第1の実施形態において説明したメモリ1122に格納される各プログラムが格納される。

【 0 1 3 8 】

本実施形態のコンピュータシステムにおける処理の一連の流れを以下に示す。

まずホスト105は、記憶装置104aに保持されるマスタデータヘリード要求を発行する。

【 0 1 3 9 】

リード要求を受信したコピー管理スイッチ503のプロトコル変換部5032は、ポートプロセッサ5033においてアドレス変換処理プログラム126を実行し、じ受信したリード要求のフレームをコピーデータを格納した記憶領域へのリード要求のフレームに変換する。

【 0 1 4 0 】

その後、プロトコル変換部5032は変換したリード要求の内容を確認する。本実施形態の場合、コピー管理スイッチ503の提供する記憶領域へのリード要求であるので、プロトコル変換部5032は、変換したリード要求のフレームをディスク管理部5035に対して送信する。

【 0 1 4 1 】

スイッチ部5034を介してリード要求を受取ったディスク管理部5035は、受信したリード要求に従い、ハードディスク5036から指定されたデータを読み出し、読み出したリードデータをスイッチ部5034を介してプロトコル変換部5032へ送信する。

【 0 1 4 2 】

リードデータを受信したプロトコル変換部5032は、ポートプロセッサ5033でアドレス変換処理プログラム126を実行し、リードデータの送信元を記憶装置104a

のSANアドレスに変更し、その変更後のリードデータのフレームをポート5031を介してホスト105へ転送する。ホスト105は、リードデータを記憶装置104aから送信されたものとして受信する。

【0143】

このようにして、ホスト105から記憶装置104aに格納されたマスタデータへのリード要求は、コピー管理スイッチ503が提供する記憶領域に格納されたコピーデータを用いてコピー管理スイッチ503によって処理される。

【0144】

本実施形態では、コピー管理スイッチ503の提供する記憶領域をコピーデータの保持に用いた。しかし、本実施形態のコピー管理スイッチ503は、上述した第一から第四の実施形態で説明したコピー管理スイッチ103と同様の使用方法に使用することができる。

【0145】

次に、本発明の第六の実施形態について説明する。上述した第1～5の実施形態では、マスタデータの全体又は決められた一部をコピーしていた。また、データのコピーは、管理者等の指示に基づいて、コピー管理テーブル131、231の設定と共に実行されていた。しかし、ここで説明する第6の実施形態では、後述するコピー管理スイッチ703が、ホスト105からのリード要求に応じて、マスタデータを格納した記憶装置104aから指定された記憶装置104bにリード要求の対象となったデータをコピーする。このように動作することで、記憶装置104bが有する記憶容量を効率よく使用出来る。

【0146】

第六の実施形態のコンピュータシステムの構成は、コピー管理スイッチ703がコピー管理スイッチ103と置き換わった以外は第1の実施形態と同じである。

【0147】

本実施形態では、コピー管理スイッチ703のメモリ1122、1142に格納される情報等が、第一の実施形態と以下の点で異なる。

第一に、CPU1121が初期化処理プログラム721を実行して、後述するコピー管理テーブル731の各コピー管理エントリ732のキャッシュインデックス737に対応す

るキャッシュテーブル741を作成する点である。

【0148】

第二に、CPU1121が端末からの要求処理プログラム722を実行して、コピー管理テーブル731の内容の変更に伴うキャッシュインデックス737の追加・削除に応じて、キャッシュテーブル741を追加・削除する処理を行う点である。

第三に、CPU1141がアドレス変換処理プログラム726を実行して、ポート111がフレームを受信した際に、リード要求に含まれるアドレス情報を判定してキャッシュ処理プログラム727の実行を制御部112に指示する点である。さらにCPU1141は、リードデータなどのアドレス情報の変換を行う。処理手順の詳細は後述する。

【0149】

第四に、CPU1121がキャッシュ処理プログラム727を実行し、キャッシュテーブル741を用いて、マスタデータへのリード要求に対応するコピーデータが存在するかどうかを判定し、存在するならばコピーデータが格納された記憶装置104（ここでは記憶装置104b）へリード要求を発行する。一方、コピーデータが存在しないならば、CPU1121は、リード要求が指定するマスタデータを記憶装置104aから記憶装置104bにコピーし、コピーデータをホスト105に送信する点である。処理手順の詳細は後述する。

【0150】

図16は、コピー管理テーブル731の構造を示す図である。コピー管理テーブル731は複数のコピー管理エントリ732を有する。各コピー管理エントリ732は、第一の実施形態とは異なり、マスタデータの格納位置とコピー先の格納位置の対応関係をSANアドレス及びLUNで管理する。具体的には、各コピー管理エントリ732は、マスタSANアドレスを登録するエントリ733、マスタLUNを登録するエントリ734、コピーSANアドレスを登録するエントリ735、コピーLUNを登録するエントリ736及びキャッシュインデックスの情報を登録するエントリ737から構成される。つまり、コピー管理エントリ732は、マスタデータが格納された記憶装置104a及びそのLUN並びにコピーデータの格納に用いる記憶装置104b及びそのLUNを指定する。

尚、キャッシュインデックスは、後述するキャッシュテーブル741を指定する為の情報である。

【0 1 5 1】

図17は、キャッシュテーブル741の構造の一例を示す図である。キャッシュテーブル741は、コピー管理テーブル731で使用されているコピー管理エントリ732の数だけ存在し、エントリ737に登録されるキャッシュインデックスによってコピー管理テーブル731と関連付けられている。各キャッシュテーブル741は複数のキャッシュエントリ742を有する。各キャッシュテーブル741のキャッシュエントリ742の数はメモリ1122の容量などから設計時に決定されている。

【0 1 5 2】

キャッシュエントリ742は、マスタLBAに登録するエントリ743、コピーLBAに登録するエントリ744、マスタLengthに登録するエントリ745、非アクセスカウンタに登録するエントリ746及び有効フラグに登録するエントリ747を有する。非アクセスカウンタはキャッシュ処理プログラム727で用いられるカウンタであり、その使われ方は後述する。有効フラグはその有効フラグが所属するキャッシュエントリ742が有効か無効かを示し、例えば1ならば有効で0なら無効であることを示す。

【0 1 5 3】

つまり有効な（エントリ747に1が登録されている）キャッシュエントリ742は、マスタLBAから始まり長さがマスタLengthである記憶装置104aの記憶領域が、コピーLBAから始まる記憶装置104bの記憶領域にコピーされている事を示す。

【0 1 5 4】

図18は、本実施形態におけるアドレス変換部114で実行される処理手順を示すフローチャートである。尚、本処理は、CPU1141がアドレス変換処理プログラム726を実行することによって行われる。

まず、ポート111でフレームを受け取ると、CPU1141は、アドレス変換処理プログラム726の実行を開始する（ステップ751）。

【0 1 5 5】

次に、CPU1141は、受信したフレームの内容がリード要求か判定する（ステッ

プ752)。受信したフレームの内容がリード要求である場合、CPU1141は、リード要求で要求されるデータが、コピー対象（すなわちマスタデータ）となっている記憶装置104のLUNに格納されたデータであるかを判定する。具体的には、CPU1141は、リード要求の送信先SANアドレス及びLUNが、エントリ733及び734に登録された値と一致するコピー管理エントリ732を探す（ステップ753）。

【0 1 5 6】

ステップ753で条件を満たすコピー管理エントリ732が見つかった場合、CPU1141は、制御部112に対してキャッシュ処理プログラム727の実行を指示する。この際、CPU1141は、受信したフレームとステップ753で選択したコピー管理エントリ732に登録された情報を制御部112に送信する（ステップ754）。ステップ754での処理の終了後、CPU1141は処理を終了する。

【0 1 5 7】

一方、ステップ752で受信したフレームの内容がリード要求ではないと判断された場合、CPU1141は、フレームがリードデータ、またはレスポンスであるか判定する（ステップ755）。フレームの内容がリードデータ又はレスポンスであった場合、CPU1141は、受信したフレームがコピーデータが格納された記憶装置104（ここでは104b）から送信されたものか判定する。具体的には、フレームの送信元がエントリ735に登録された値と一致するコピー管理エントリ732がコピー管理テーブル731に存在するか判定する（ステップ756）。

【0 1 5 8】

ステップ756で条件を満たすコピー管理エントリ732がコピー管理テーブル731に存在する場合、CPU1141は、ステップ756で見つけたコピー管理エントリ732を用いて、フレームの送信元をエントリ733に登録されたマスタSANアドレスに変更する（ステップ757）。

【0 1 5 9】

ステップ753でリード要求であるフレームの送信先がコピー管理テーブル731に存在しない場合、ステップ755でフレームがリードデータでもレスポンスでもないと判断された場合及びステップ756でフレームの送信元がコピー管理テーブル731に存在しなかった場合には、CPU1141は、スイッチング処理部115に処理したフ

レームを転送して処理を終了する（ステップ758）。

【0160】

図19は、ステップ754におけるアドレス変換部114の指示によって制御部112で実行される処理手順を説明する図である。本処理は、制御部112がキャッシュ処理プログラム727を実行することで行われる。

【0161】

アドレス変換部114の指示を受けた制御部112は、アドレス変換部114から受信したフレームの内容であるリード要求のLBA及びLengthで指定されるマスターデータを格納する記憶装置104aの記憶領域が、コピーデータが格納される記憶装置104bに格納されているかどうかを判定する。具体的には、CPU1121は、アドレス変換部114から転送されたコピー管理エントリ732のエントリ737に登録された値で指定されるキャッシュテーブル741に、リード要求のLBA及びLengthで指定される領域がエントリ743及びエントリ745で指定される領域に含まれ、かつ有効フラグが1であるキャッシュエントリ742が存在するか判定する（ステップ761）。

【0162】

ステップ761の条件を満たすキャッシュエントリ742が存在する場合、CPU1121は、ステップ761で使用されたキャッシュテーブル741に含まれる有効なキャッシュエントリ742の非アクセスカウンタ746を操作する。具体的には、CPU1121は、ステップ761の条件を満たすキャッシュエントリ742のエントリ746の値を0にセットし、他の全ての有効なキャッシュエントリ742のエントリ746に登録された値を1増やす（ステップ762）。

【0163】

その後、CPU1121は、コピーデータへのリード要求を生成する。具体的には、CPU1121は、アドレス変換部114から送信されたコピー管理エントリ732のエントリ735及び736に登録された値を各々送信先SANアドレスとLUNとする。更に、CPU1121は、ステップ761で見つかったキャッシュエントリ742に登録された情報を用いて、LBAを（リード要求で指定されるLBA+エントリ744に登録された値-エントリ743に登録された値）に、Lengthをエントリ745に登録された値にする（ステップ763）。その後、CPU1121は、スイッチング処理部115に処理したフレームを送信す

る（ステップ764）。

【0 1 6 4】

一方、ステップ761の条件を満たすキャッシュエントリ742が存在しない場合、CPU1121は、コピー管理エントリ732が指定するコピーデータを格納する記憶装置104bの記憶領域にリード要求が指定するLengthのデータを格納できる分の空き領域が存在するかを判定する。また、CPU1121は、キャッシュテーブル741に、使われていないキャッシュエントリ742が存在するかを判定する。具体的には、CPU1121は、キャッシュテーブル741の有効な全てのキャッシュエントリ742の情報を用いて、記憶装置104bで使用されている記憶領域を判別し、それ以外の記憶領域でリード要求に含まれるLengthの値以上連続している記憶領域を記憶装置104bから探す（ステップ765）。

【0 1 6 5】

有効ないずれのキャッシュエントリ742にも使用されておらず、リード要求のLength以上の長さの記憶領域が記憶装置104bに見つからない場合、もしくは全てのキャッシュエントリ742が有効なため、キャッシュエントリ742の更新が行えない場合、CPU1121は、キャッシュテーブル741が有する有効なキャッシュエントリ742のうち1つを削除すなわち無効化して記憶領域の空き領域を広げる。具体的には、CPU1121は、有効なキャッシュエントリ742の中で最もエントリ746に登録された値が大きいキャッシュエントリ742のエントリ747の値を0にする（ステップ766）。その後、CPU1121は、ステップ765の処理を繰り返す。

【0 1 6 6】

有効ないずれのキャッシュエントリ742にも使用されておらず、リード要求のLength以上の長さの記憶領域が記憶装置104bに見つかり、さらに無効なキャッシュエントリ742が1つでもある場合、見つかったキャッシュエントリ742を更新し、リード要求が指定するマスターデータと、見つかった記憶装置104bの記憶領域の対応を登録する。そのため、まずマスターデータのコピーを行う。はじめに該当する記憶領域が見つかり、かつ無効なキャッシュエントリ742が1つでも見つかった場合、CPU1121は、リード要求で指定されるマスターデータを、マスターデータを保持する記憶装置104aから読み出す。具体的には、送信元を制御部112のSANアド

レスにしたマスタデータのリード要求を作成し、スイッチング処理部115に送信する。その後、CPU1121は、マスタデータを保持する記憶装置から送られてくるリードデータを、メモリ1122に格納する（ステップ767）。

【0 1 6 7】

その後、CPU1121は、メモリ1122に格納されたマスタデータを、ステップ765で見つけた記憶装置104bの記憶領域に転送する（ステップ768）。その後、CPU1121は、キャッシュテーブル741を更新する。具体的には、CPU1121は、ステップ765で見つかった無効なキャッシュエントリ742について、メモリ1122から転送されたデータが格納された記憶領域に対応するコピーLBAをエントリ744に、そのマスタデータの記憶領域に対応するマスタLBA及びマスタLengthをエントリ743及び745に登録し、かつそのキャッシュエントリ742のエントリ746の値を0、エントリ747の値を1にセットする（ステップ769）。その後、CPU1121は、ステップ763以降の処理を行う。

【0 1 6 8】

本実施形態の一連の動作をまとめると、次のようになる。

【0 1 6 9】

はじめに、システムの利用者や管理者が、コピー管理スイッチ703に接続された管理端末106を介して、記憶装置104aのあるLUNで指定される記憶領域に格納されたマスタデータを記憶装置104bのあるLUNで指定される記憶領域にコピーする設定をコピー管理テーブル731に行う。そして、この時点では、データは記憶装置104bにコピーされていない状態とする。つまり、コピー管理テーブル731のコピー管理エントリ732は存在するが、そのコピー管理エントリ732のキャッシュインデックス737に対応するキャッシュテーブル741のキャッシュエントリ742は全て無効になっている。これが初期状態となる。

【0 1 7 0】

初期状態においてホスト105が記憶装置104aに保持されているマスタデータへのリード要求を発行すると、リード要求のフレームがコピー管理スイッチ703に到達する。

【0 1 7 1】

フレームの受信を確認すると、アドレス変換部114はアドレス変換処理プログラム726を起動する。ここでは、アドレス変換部114は、リード要求がマスターデータへの要求であると判定し、制御部112へキャッシュ処理プログラム727の実行を指示する。

【 0 1 7 2 】

指示を受けた制御部112は、リード要求で指定されるデータに対応するキャッシュエントリ742を探す。この時点でデータは一切記憶装置104bに転送されていないので、キャッシュエントリ742は見つからない。そこで制御部112は、記憶装置104aに対してリード要求を発行し、そのリードデータを記憶装置104bに転送する。また制御部112は、記憶装置104bに転送されたデータに対応するキャッシュエントリ742をキャッシュテーブル741に登録する。

【 0 1 7 3 】

更に、制御部112は、受信したリード要求のフレームからホスト105を送信元にしたコピーデータへのリード要求のフレームを生成し、そのフレームをスイッチング処理部115に転送する。

【 0 1 7 4 】

スイッチング処理部115はフレームを記憶装置104bに転送する。フレームを受取った記憶装置104bは、ホスト105を送信先としたリードデータのフレームをコピー管理スイッチ703に経由でホスト105へ送信する。

【 0 1 7 5 】

リードデータのフレームを受取ったアドレス変換部114は、この場合、受信したフレームのリードデータがコピーデータだと判定する。アドレス変換部114は、リードデータの送信元を記憶装置104aに変更し、スイッチング処理部115に転送する。

スイッチング処理部115はリードデータをホスト105に転送する。リードデータは記憶装置104aから送信されたデータとしてホスト105に到達する。

【 0 1 7 6 】

以上の流れにより、ホスト105からのリード要求によって記憶装置104bにコピーが作成され、キャッシュエントリ742がキャッシュテーブル741に作成される。

【0177】

その後、ホスト105が再び同じマスタデータへリード要求を発行すると、リード要求のフレームを受信した制御部112は、当該マスタデータに対応するキャッシュエントリ742を見つけることが出来る。この場合、制御部112は、リード要求のフレームに基づいてホスト105を送信元にしたコピーデータへのリード要求を生成し、スイッチング処理部115に転送する。

【0178】

このように、2度目のリード要求は記憶装置104bに格納されたコピーデータだけを用いて処理される。このためリード要求の処理が高速化される。また、本実施形態においては、実際にホストがアクセスするマスタデータのみがコピーされる。更に、非アクセスカウンタ746を用いて最も長い間アクセスされないコピーデータがキャッシュエントリ742から削除される。つまり、ホストからのアクセス頻度が高いデータのみが記憶装置104bに格納されることとなる。したがって、本実施形態では、記憶装置104bの記憶容量を効率的に使用できる。

【0179】

本実施形態でキャッシュエントリ742を削除するために用いた方法は、一般にLRU(Least Recently Used)と呼ばれる方法であるが、他の方法でも良い。

また、本実施形態では、コピー管理装置は、ホスト105からのリード要求で指定されるマスタデータを記憶装置104aから読み出し、そのリードデータをメモリ1122に保存してからコピーデータを格納する記憶装置104bに転送した。しかし、ホスト105からのリード要求が指定するマスタデータだけでなく、その前後を含むデータを記憶装置104aから読み出しても良い。

【0180】

例えば、ホスト105からのリード要求が指定する記憶領域に対して、同じ先頭位置から始まりLengthをより長くした記憶領域からデータを読み出すとしても良い。こうすると、ホスト105が連続した領域へリード要求を行う場合に、次のリード要求が記憶装置104bに存在する可能性が高くなる。また、SANアドレスとLUNで指定される記憶領域全体を読み出しても良い。

【0181】

また、記憶装置104aがSCSIのEXTENDED COPYコマンドを処理できる場合は、記憶装置104aに対するデータのリード要求を、記憶装置104aから記憶装置104bへのデータのコピー要求に置き換えても良い。さらに、マスタデータをコピーデータを格納する記憶装置104bに転送するための専用のバッファメモリを制御部112に用意しても良い。さらに、コピーを保持する記憶装置104bにリードデータを転送しながら同時にホスト105にリードデータを送信しても良い。

【 0 1 8 2 】

【発明の効果】

本発明によれば、SANに接続された記憶装置に保持されたデータへのアクセスを高速化することができる。

また本発明によればSANを流れるデータを減らすことができるので、SANの負荷を減らすことができる。

【図面の簡単な説明】

【図 1】

本発明を適用したコンピュータシステムの第一の実施形態の構成を示す図である。

【図 2】

第 1 の実施形態のコピー管理スイッチの構成を示す図である。

【図 3】

第 1 の実施形態のコピー管理スイッチのメモリ構成を示す図である。

【図 4】

第 1 の実施形態のコピー管理テーブルの構成を示す図である。

【図 5】

第 1 の実施形態のアドレス変換処理のフローチャートである。

【図 6】

本発明を適用したコンピュータシステムの第二の実施形態の構成を示す図である。

【図 7】

第 2 の実施形態の代理アドレステーブルの構成を示す図である。

【図 8】

第 2 の実施形態のアドレス変換処理のフローチャートである。

【図 9】

本発明を適用したコンピュータシステムの第三の実施形態の構成を示す図である。

【図 10】

第 3 の実施形態の仮想アドレステーブルの構成を示す図である。

【図 11】

第 3 の実施形態におけるコピー管理テーブルの例を示す図である。

【図 12】

第 3 の実施形態においてコピー管理スイッチが複数ある構成例を示す図である。

【図 13】

本発明を適用したコンピュータシステムの第 4 の実施形態の構成を示す図である。

【図 14】

本発明を適用したコンピュータシステムの第 5 の実施形態の構成を示す図である。

【図 15】

コピー管理スイッチの構成例を示す図である。

【図 16】

第 6 の実施形態のコピー管理テーブルの構成を示す図である。

【図 17】

第 6 の実施形態のキャッシュテーブルの構成を示す図である。

【図 18】

第 6 の実施形態のアドレス変換処理のフローチャートである。

【図 19】

第 6 の実施形態のキャッシュ処理のフローチャートである。

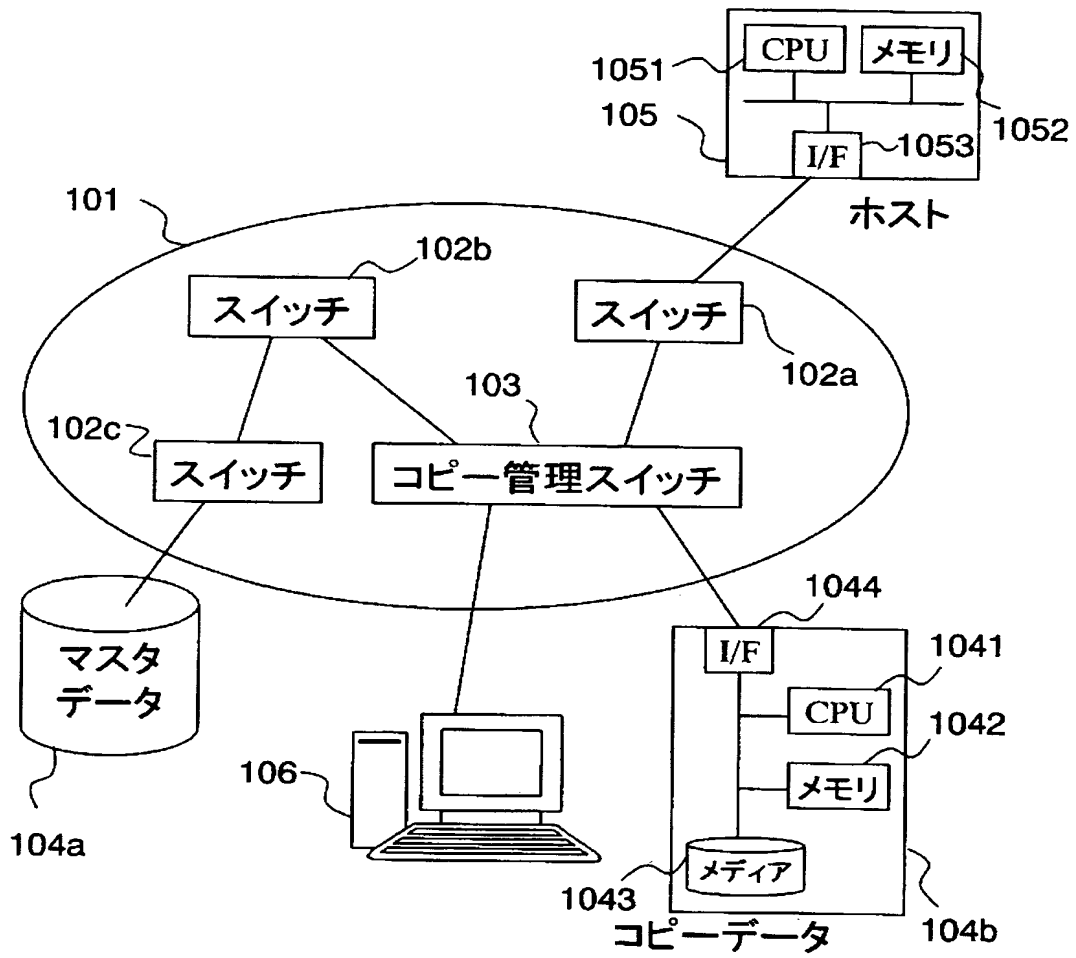
【符号の説明】

101…ストレージエリアネットワーク、102…スイッチ、103…コピー管理スイッチ、104…記憶装置、105…ホスト、106…管理端末。

【書類名】 図面

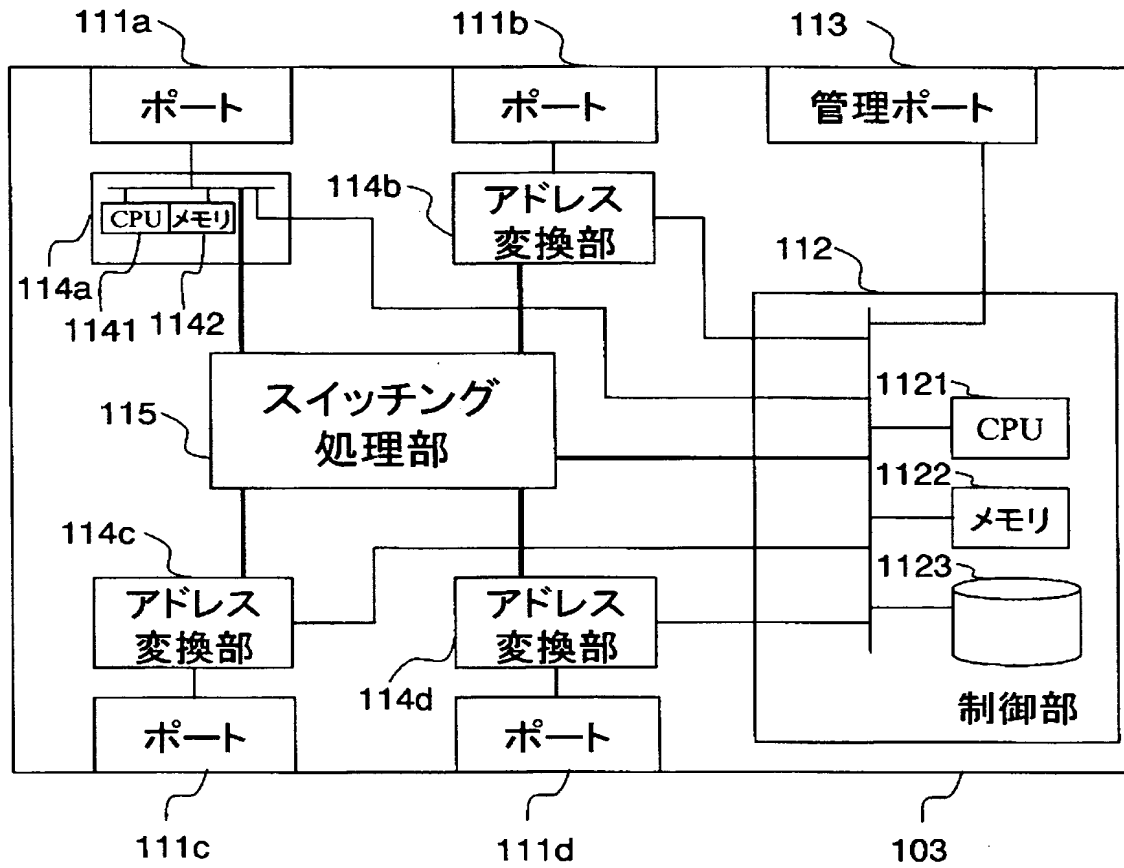
【図 1】

図 1



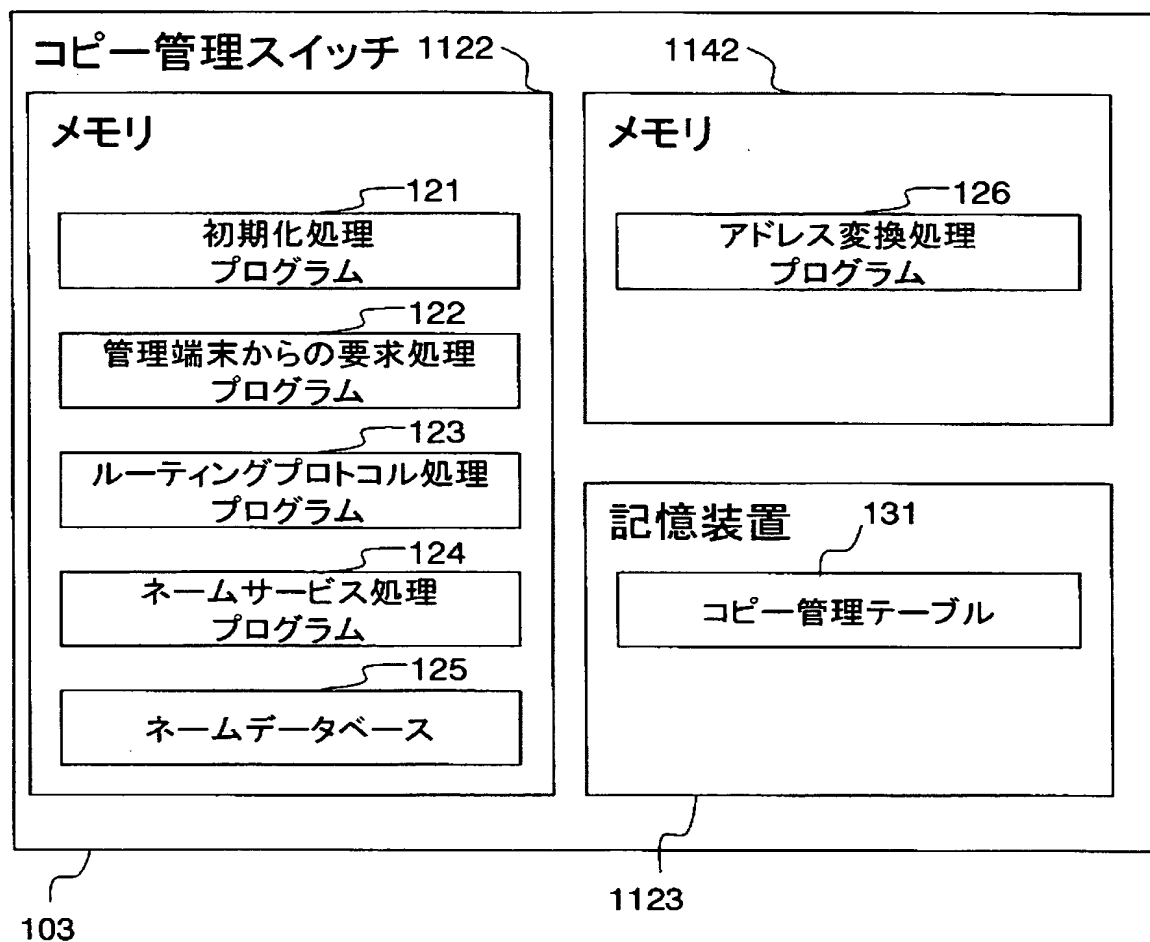
【図 2】

図 2



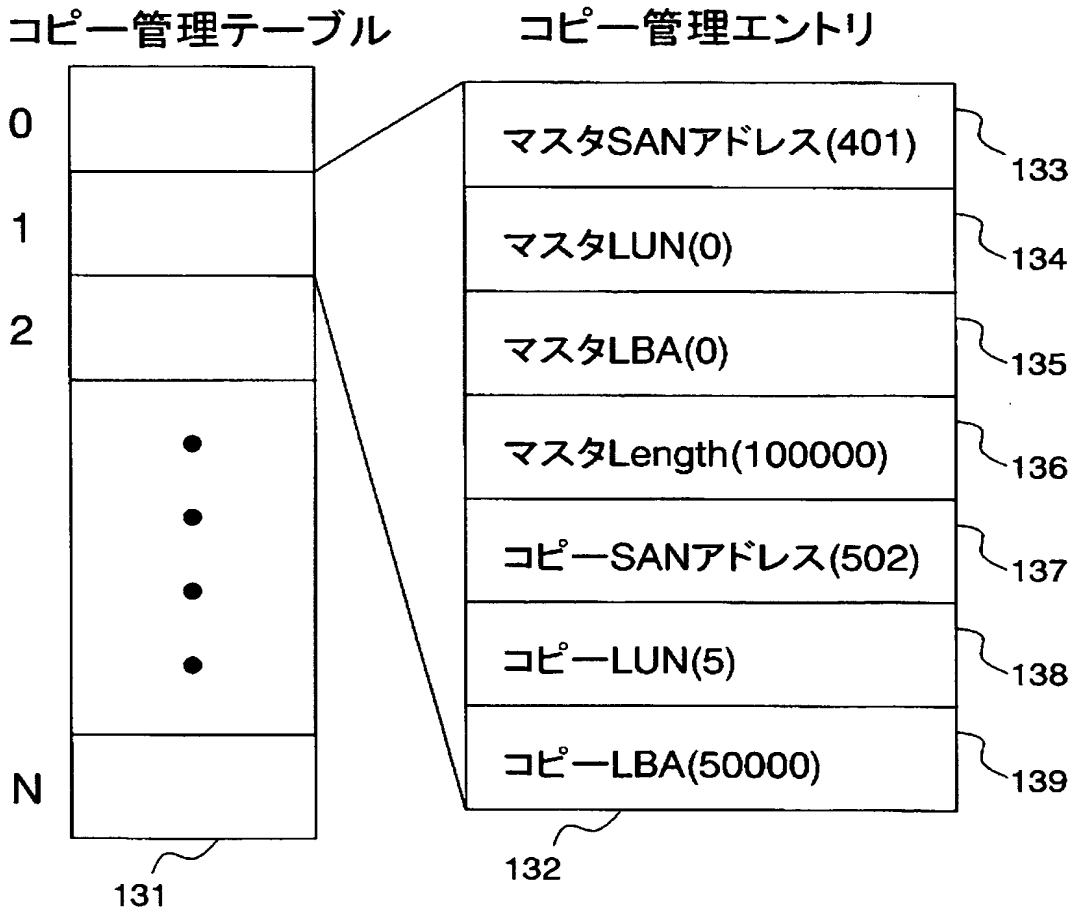
【図 3】

図 3



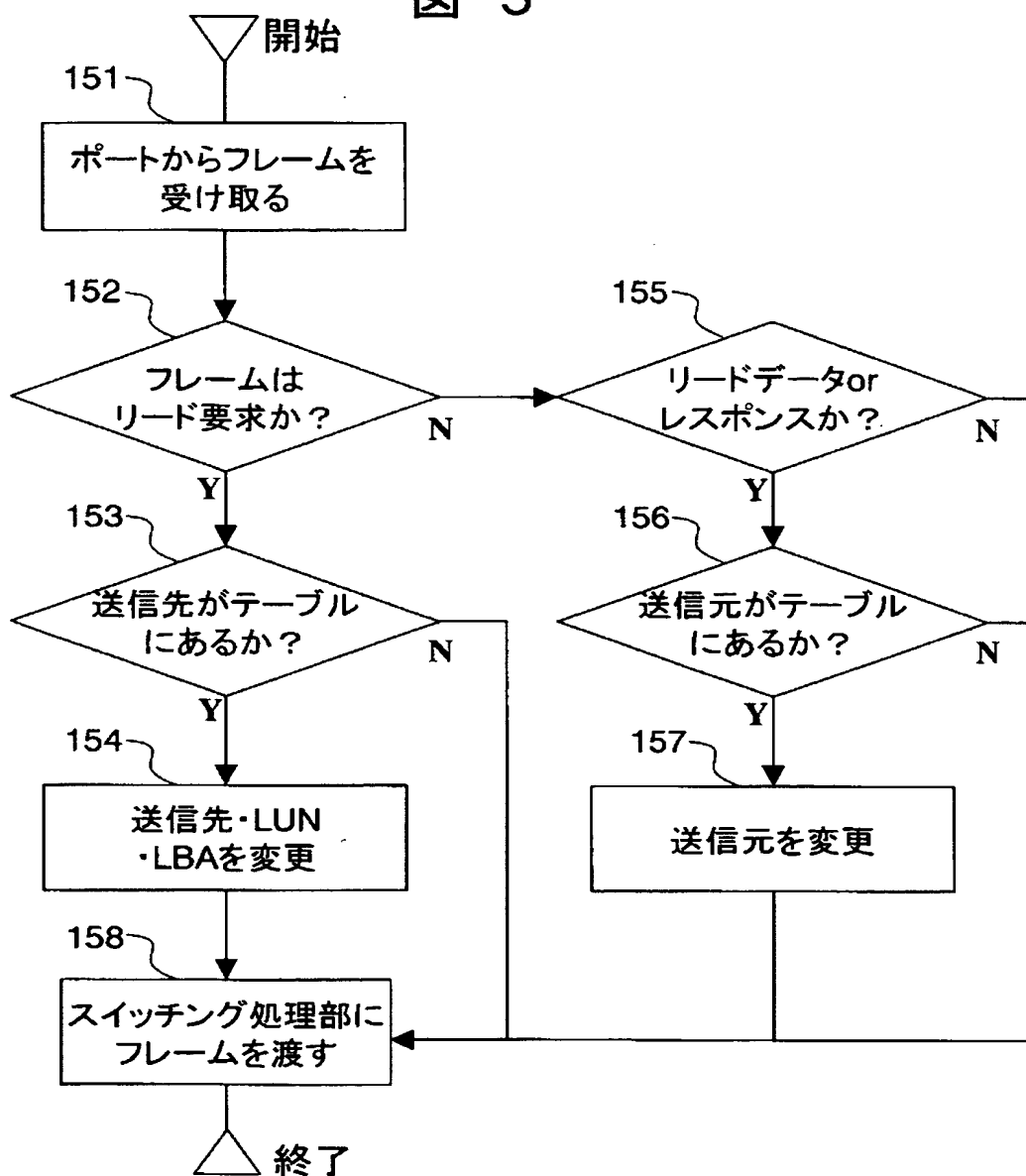
【図 4】

図 4



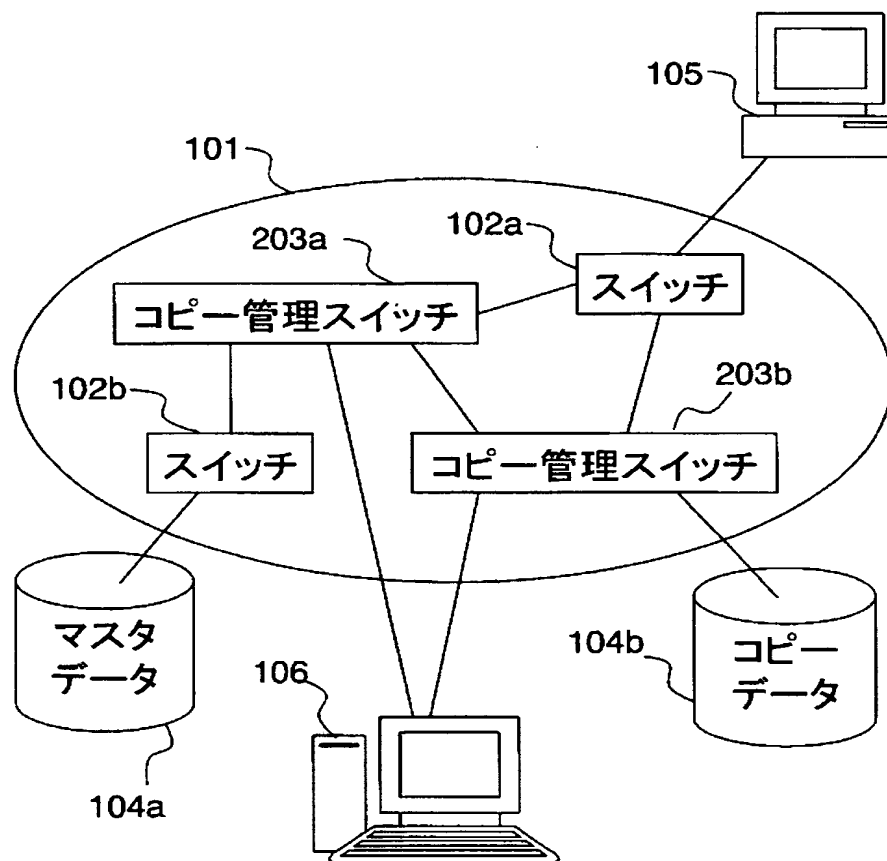
【図 5】

図 5



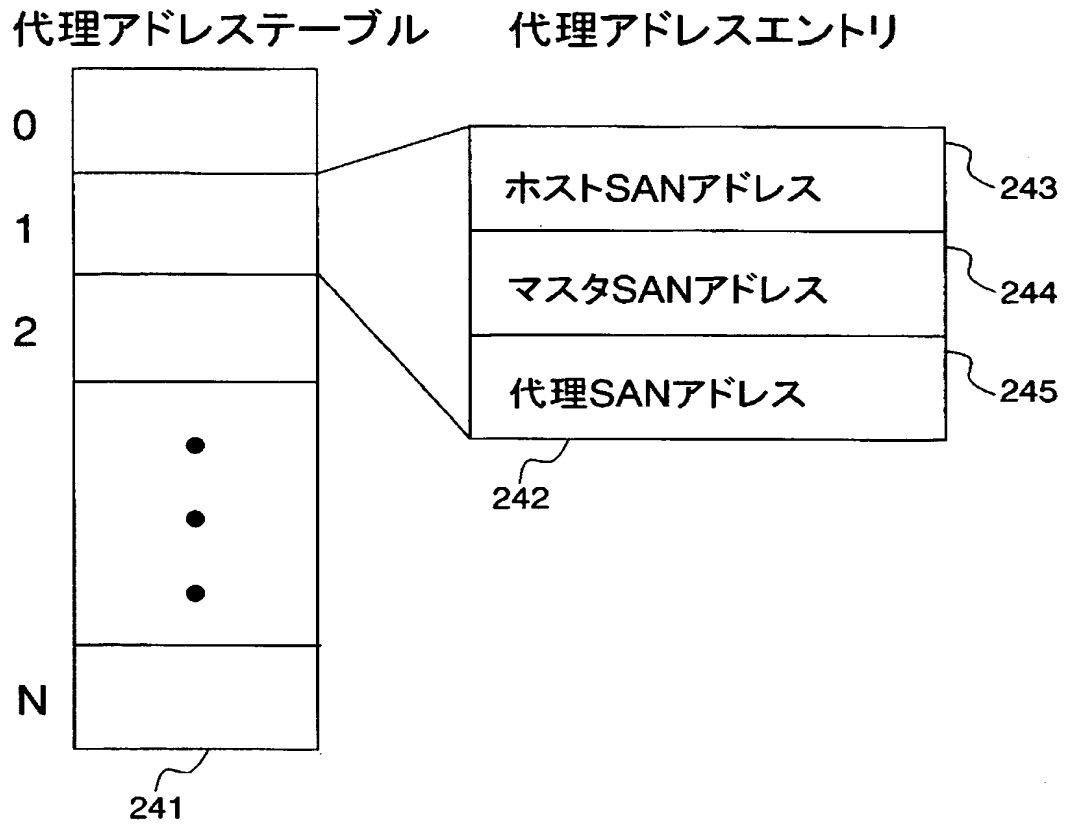
【図 6】

図 6



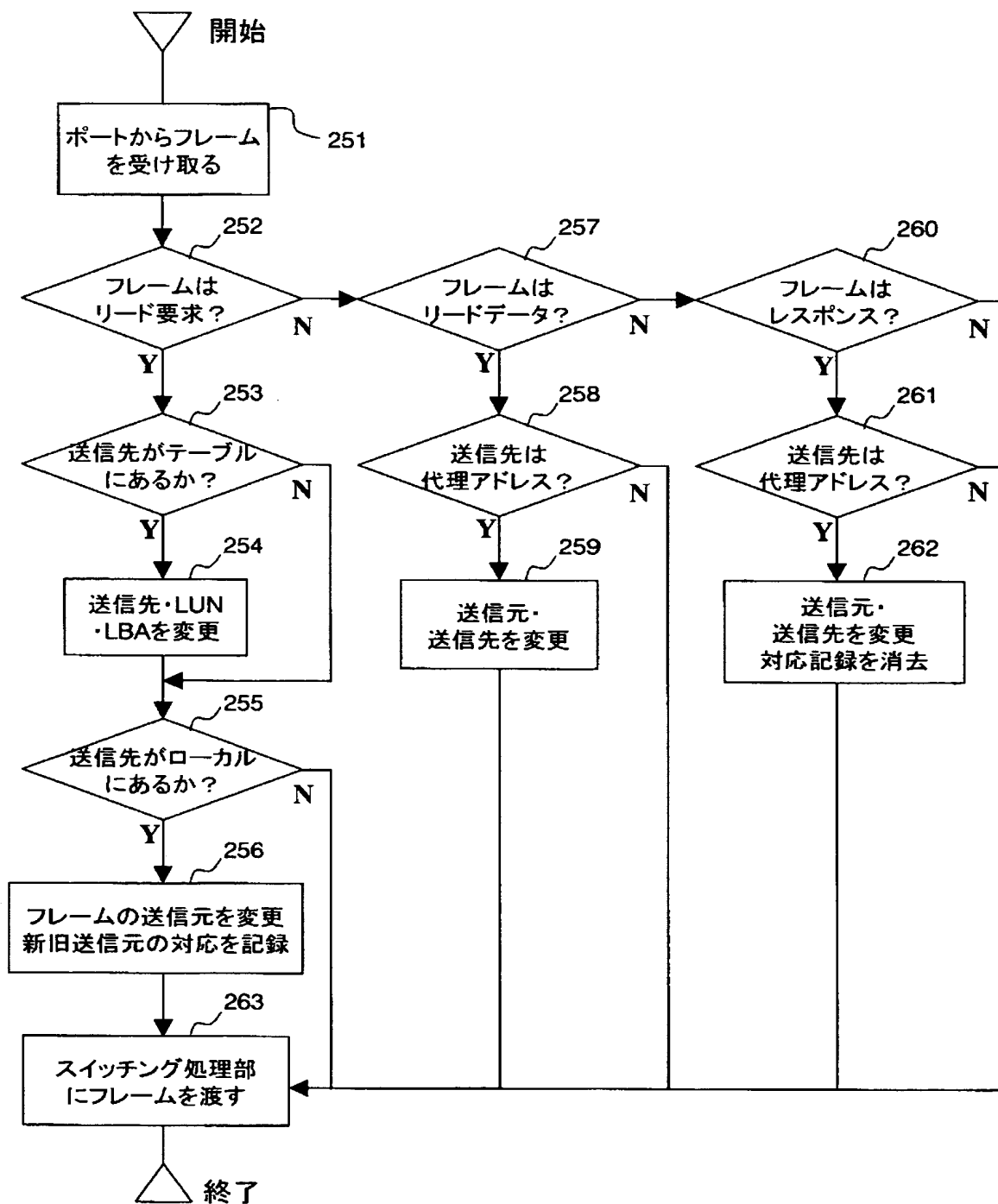
【図 7】

図 7



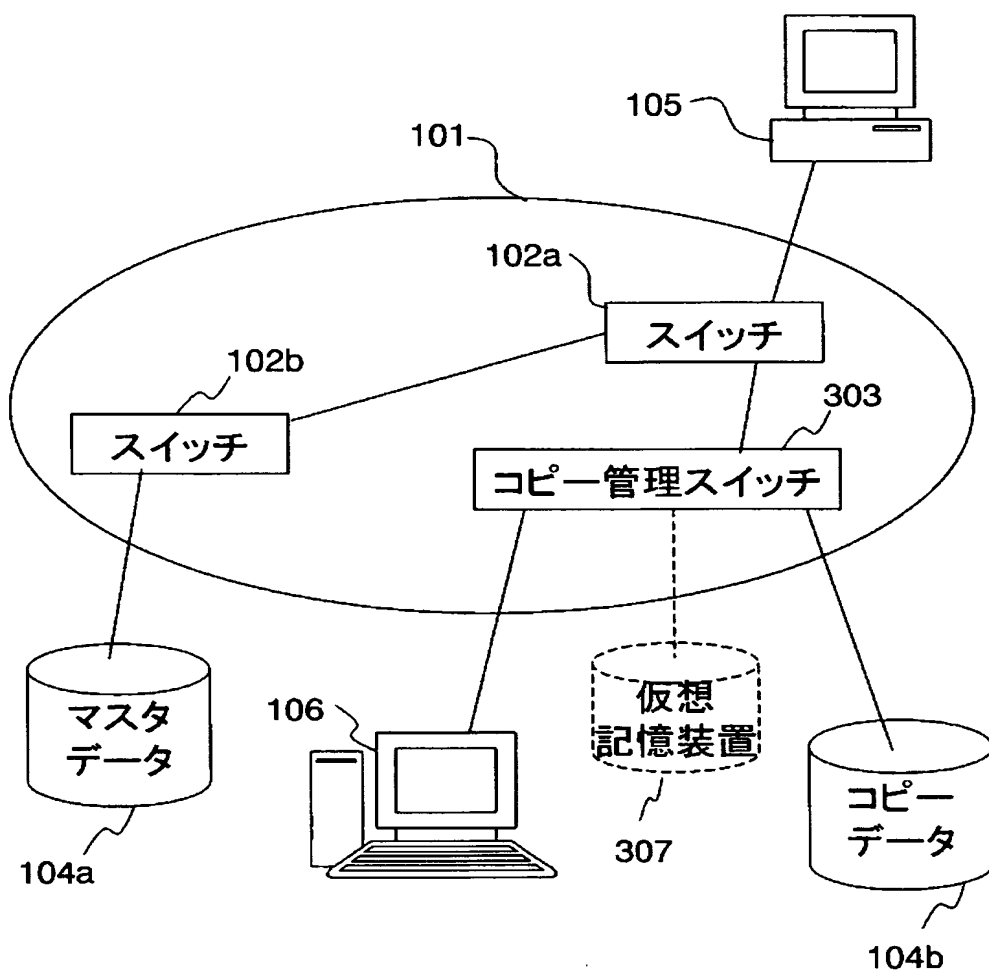
【図 8】

図 8



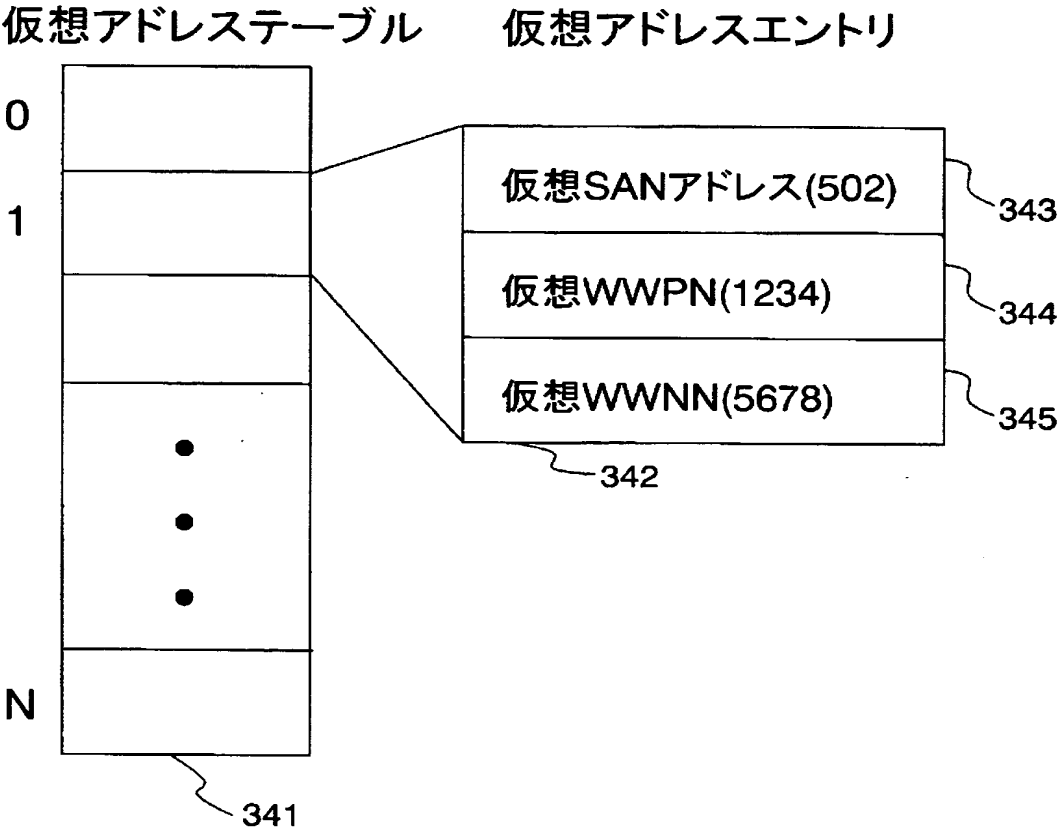
【図 9】

図 9



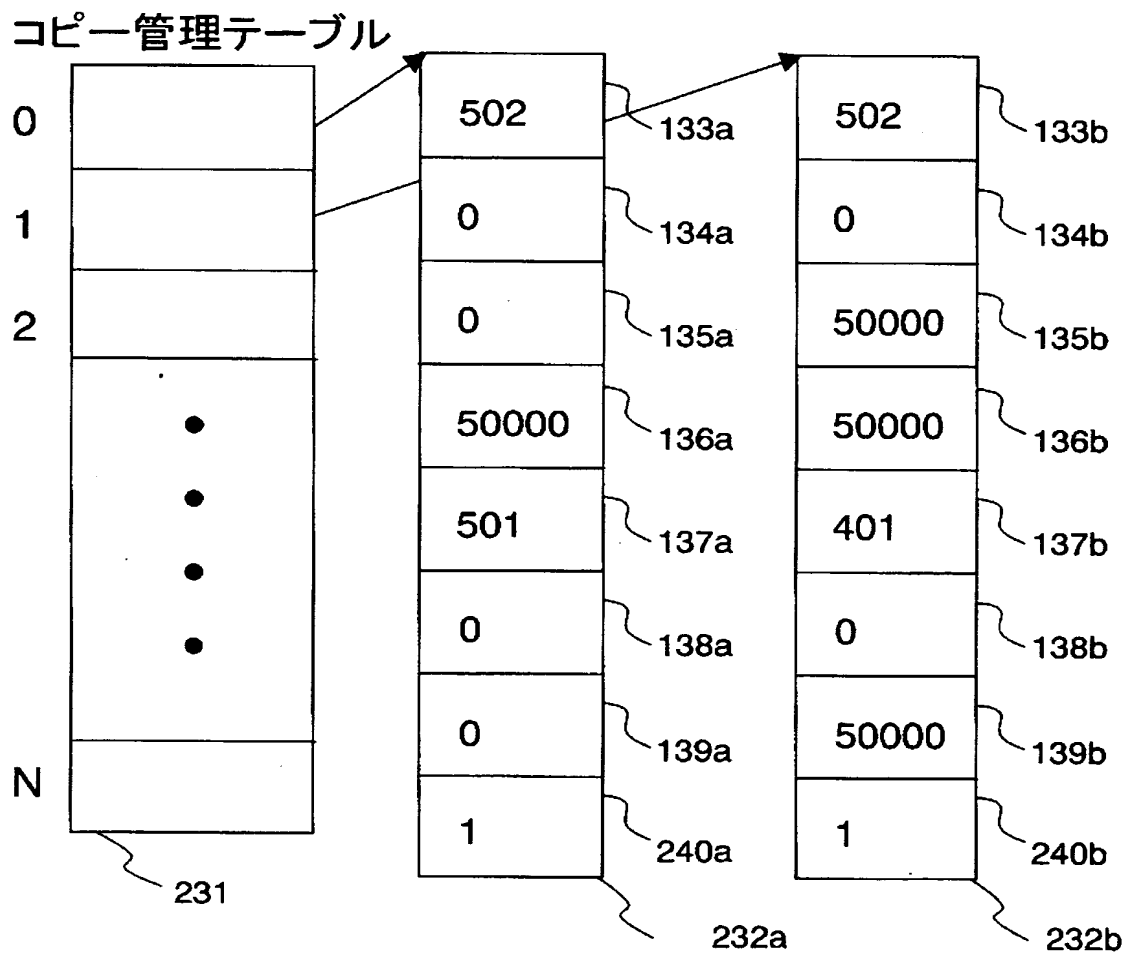
【図10】

図 10



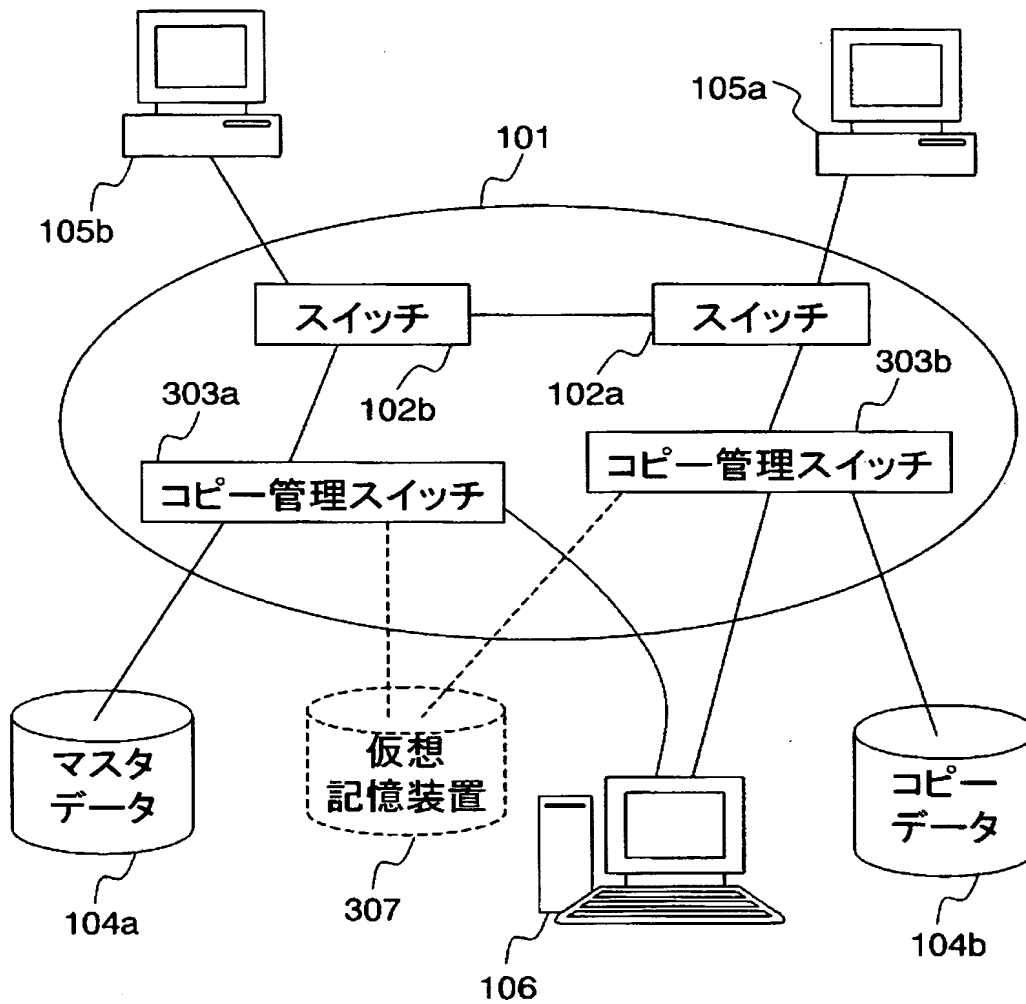
【図 11】

図 11



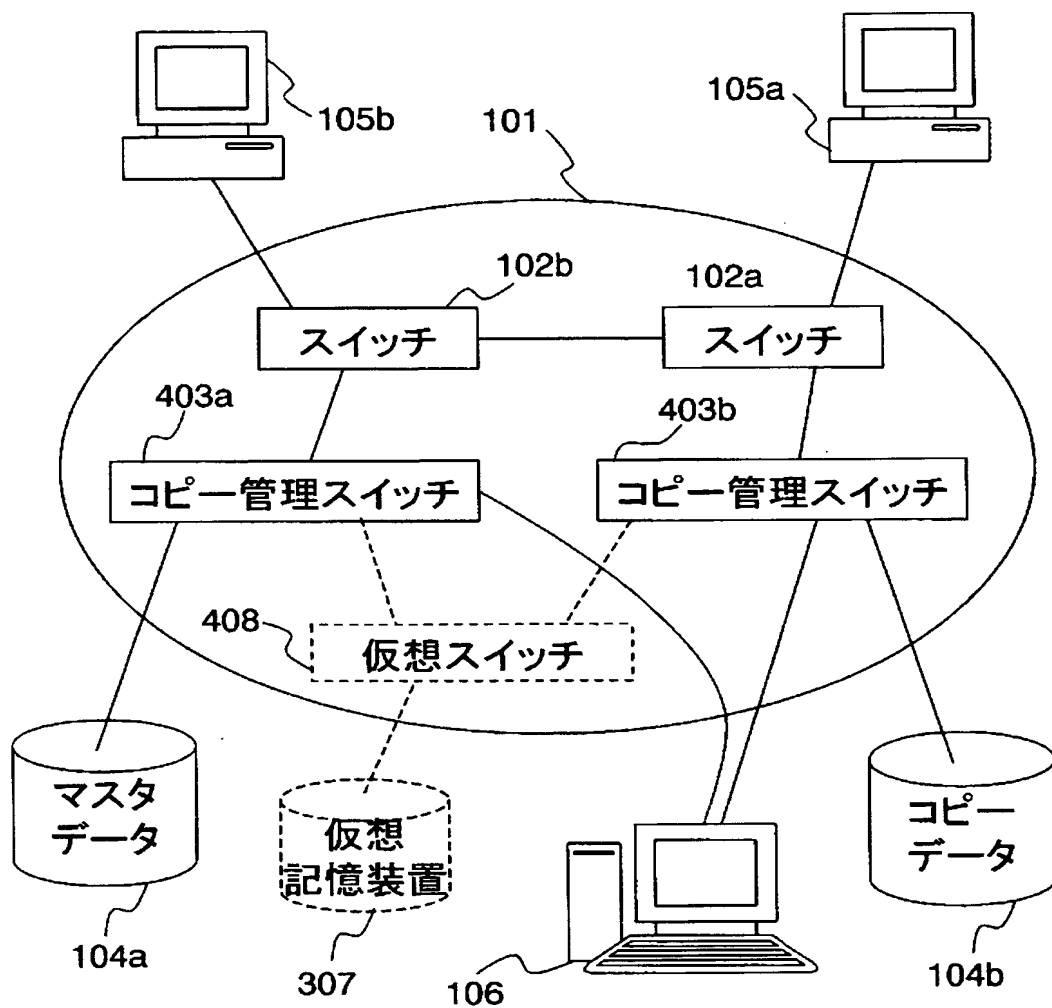
【図 12】

図 12



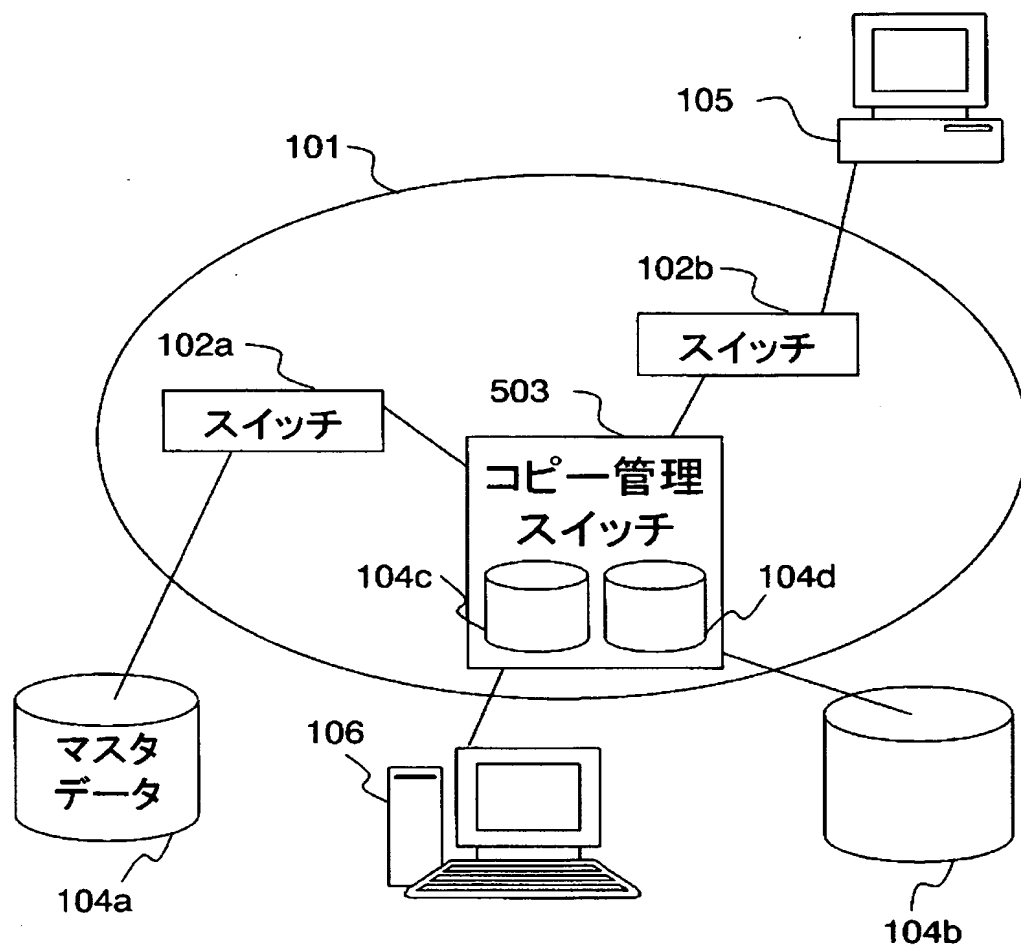
【図 13】

図 13



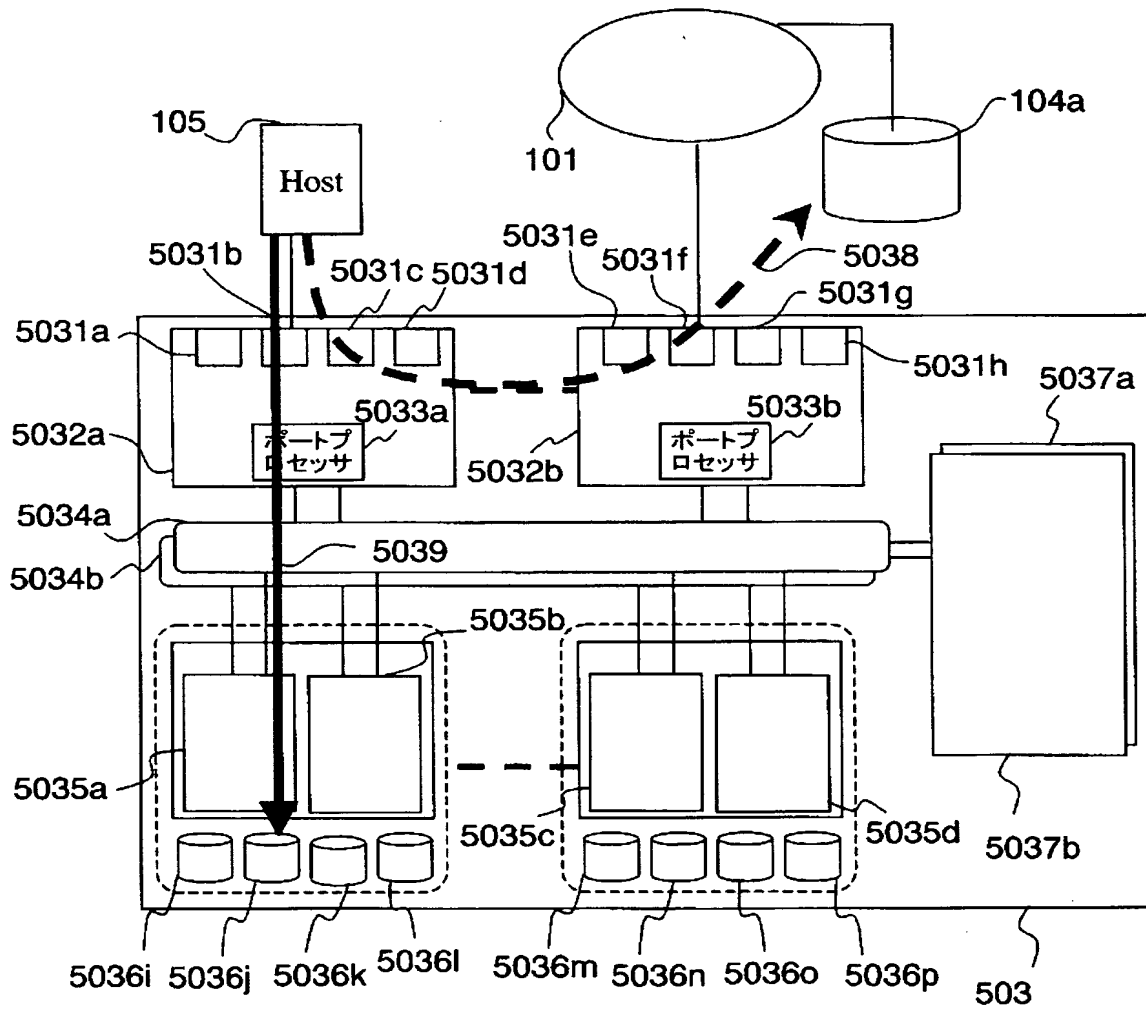
【図 14】

図 14



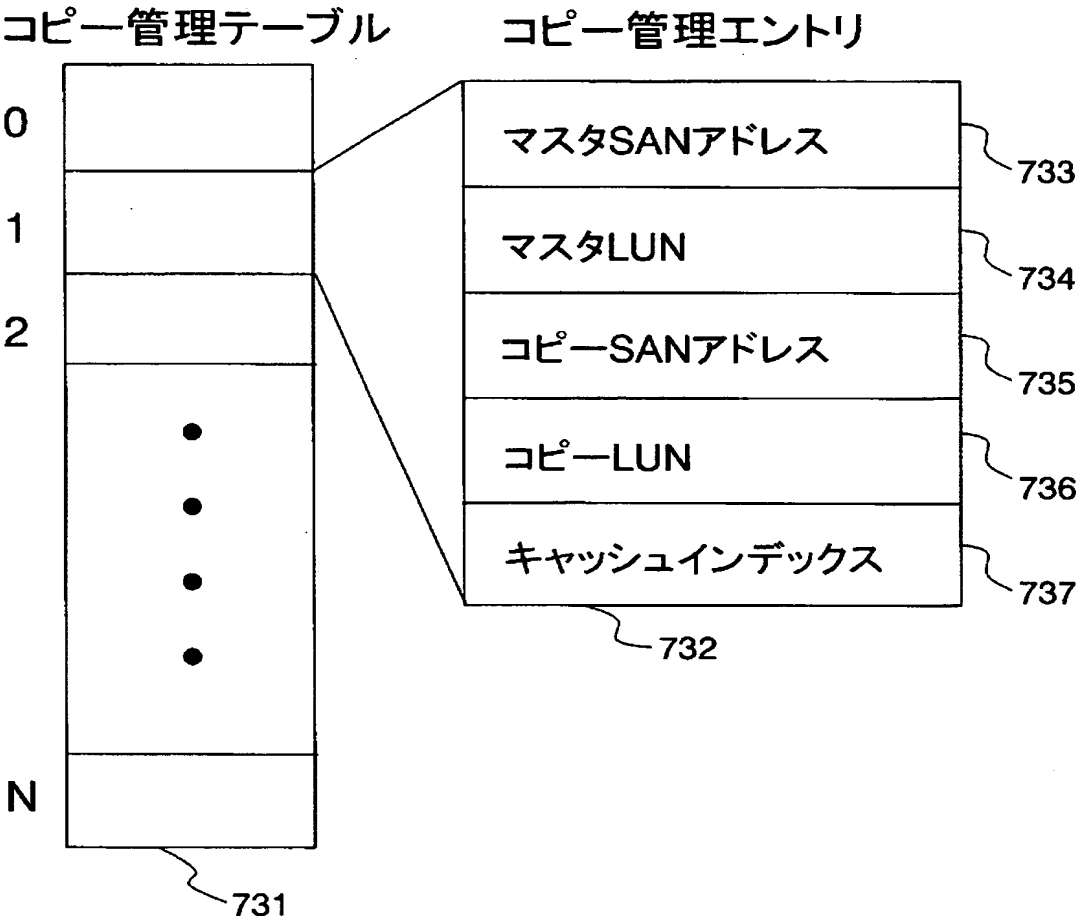
【図 15】

図 15



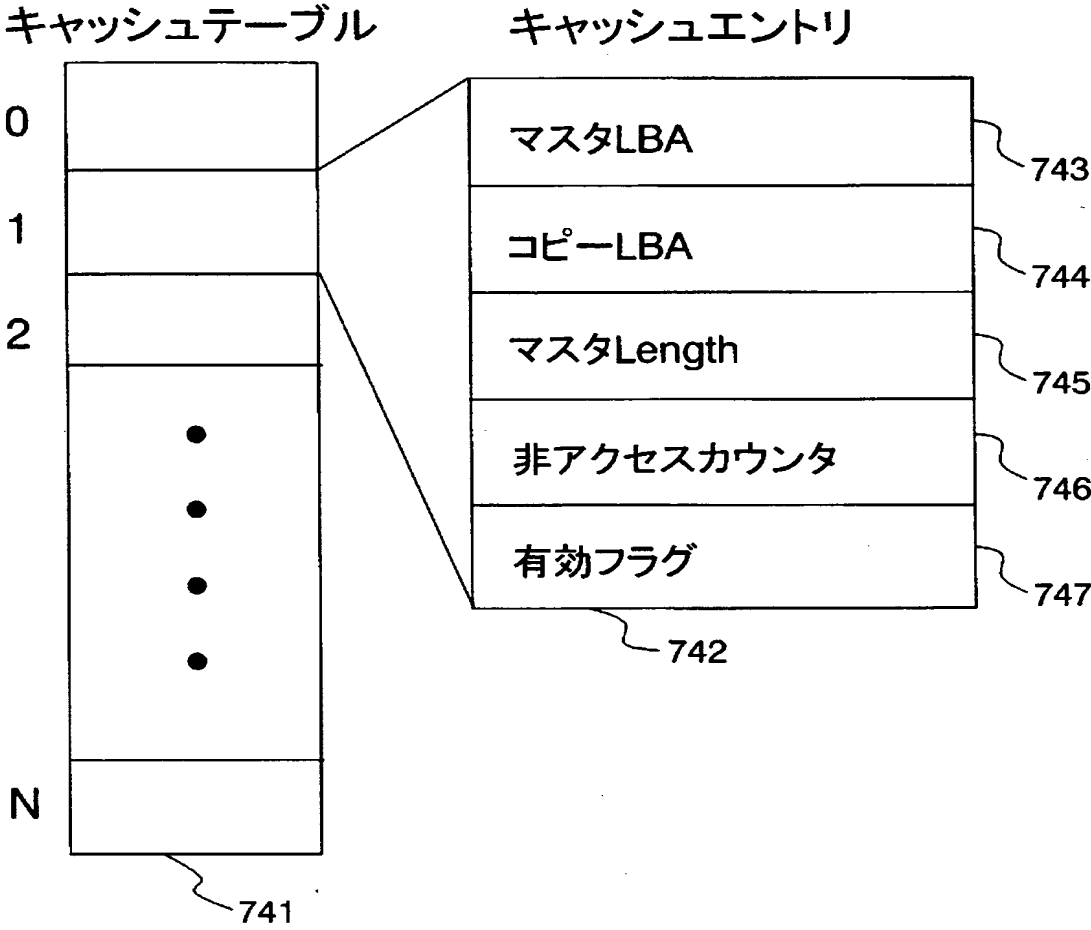
【図 16】

図 16



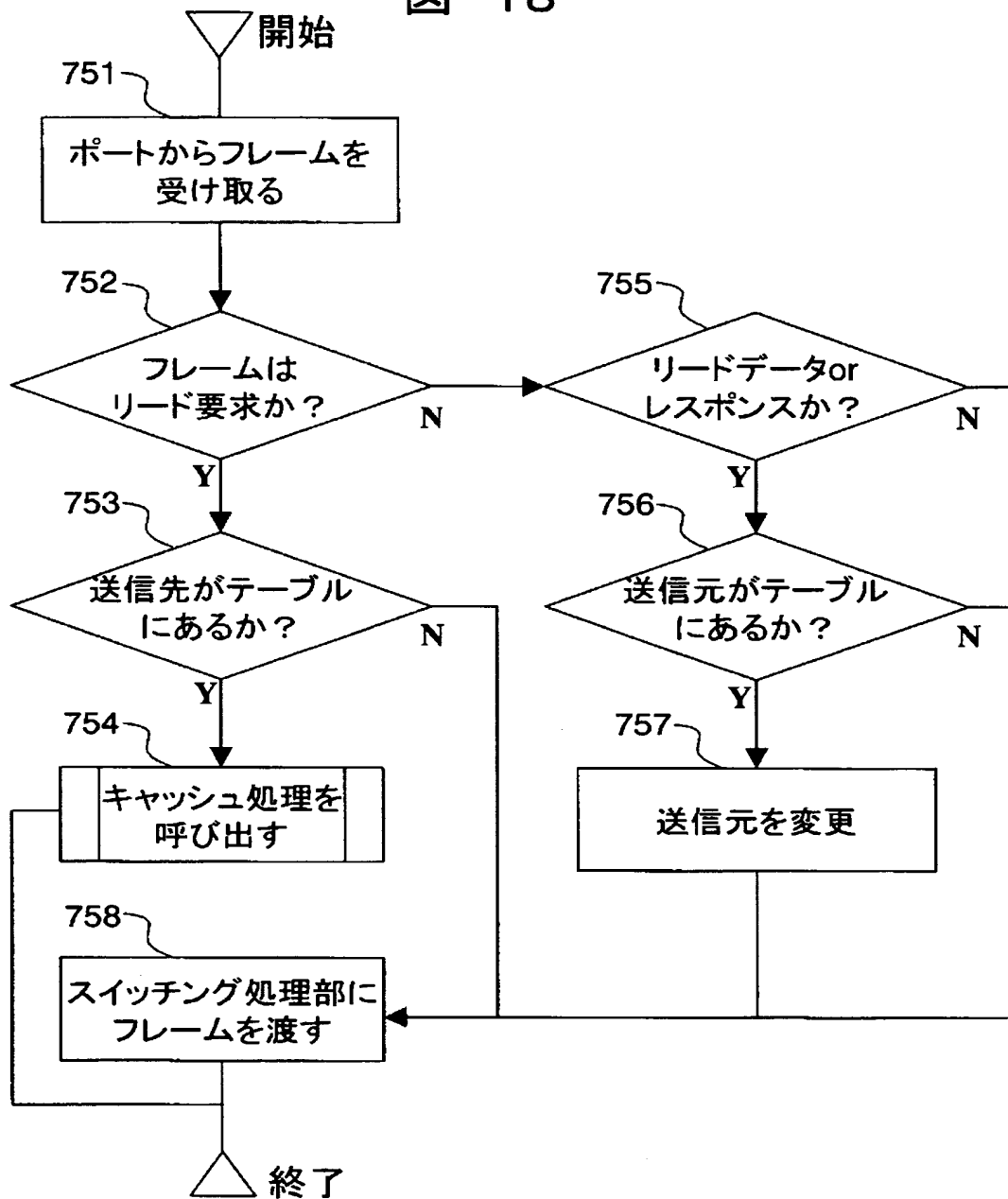
【図 17】

図 17



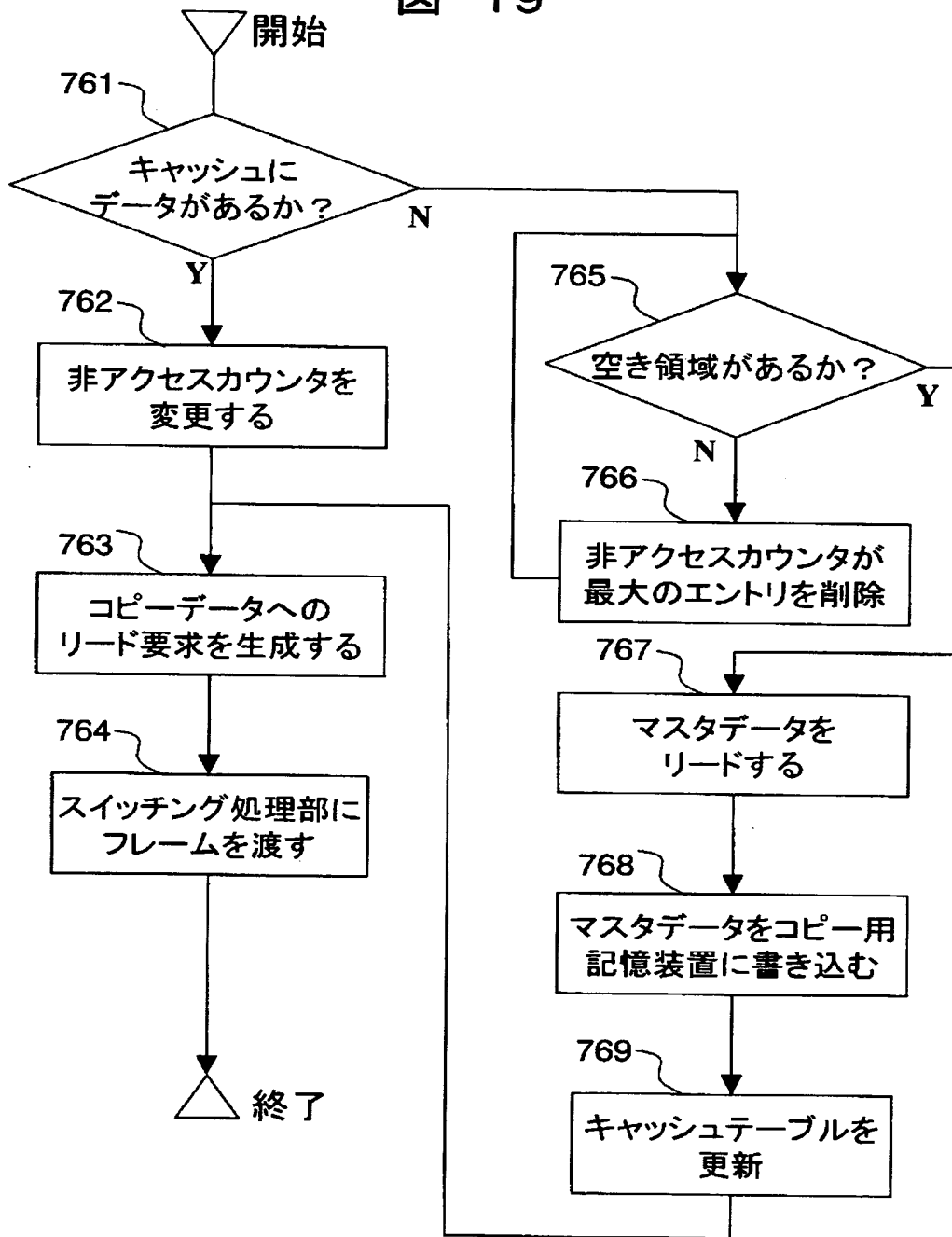
【図18】

図 18



【図 19】

図 19





【書類名】 要約書

【要約】

【課題】 S A N上でのデータへのアクセスを高速化し、ネットワークの負荷を減少させる。

【解決手段】

コピー管理スイッチがデータへのリード要求を、より近くにあるコピーデータを保持する記憶装置へのリード要求に変換する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 1 7 2 7 7 0
受付番号	5 0 3 0 1 0 1 3 8 3 1
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 6 月 1 9 日

< 認定情報・付加情報 >

【提出日】	平成15年 6月18日
-------	-------------

次頁無

特願 2 0 0 3 - 1 7 2 7 7 0

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所